

# 包囲型マイクロホンアレイを用いた室内音源特性の高精度推定法

著者	岡本 拓磨
学位授与機関	Tohoku University
URL	<a href="http://hdl.handle.net/10097/39933">http://hdl.handle.net/10097/39933</a>

東北大学審査博士学位論文

包囲型マイクロホンアレイを用いた  
室内音源特性の高精度推定法

2009年1月20日

東北大学大学院情報科学研究科

システム情報科学専攻

岡本 拓磨



# 緒言

人間は、目で物を見る、耳で音を聞く、体で物を触る、などの行動によって、その物が何であるのかを認識し、生活している。さらに、物の内容だけでなく、物までの距離や方向、物の動き、物がある空間(部屋)の大きさなどといった空間的な情報も同時に認識している。目で物を見た場合、近くにあるものは大きく、遠くにあるものは小さく見える。音を聞いた場合も、音が左で鳴っている時は音が左方向から聞こえ、遠くで音が鳴っている時は近くで鳴っている時よりも音が小さく聞こえる。

人間は普段から視覚や聴覚からこのようにして実空間に存在する物体や空間情報を認識している。人間のこれらの感覚や知覚の仕組みを利用して、視覚や聴覚への入力情報を制御することにより、人間に対して物体や空間情報を積極的に認識させることができる。絵画や実環境を記録した静止画や動画を平面で見たときでも、人間は映し出された物の配置関係などから遠近感のような空間情報を知覚できる。また、スピーカを2つ用いたステレオ再生などによって、人間は音の広がりや動きといった空間情報を知覚できる。

近年「人間に対して物体や空間情報を積極的に認識させることができる」技術が注目を集めている。人間は普段から平面(2次元)ではなく立体(3次元)の物を見ており、ステレオ(空間での2点)ではなく、あらゆる方向からの音を聞いている。よって、視覚への情報を立体映像と制御して提示すること、聴覚への情報を多方向からの音として制御して提示することにより、平面映像やステレオの音を聞いたときよりも、高い臨場感を得ることができる。このように、実際の物がそこにある、その場にいるかのような感覚を仮想的に作り出し、提示する技術をバーチャルリアリティ技術と呼ぶ。立体映像提示技術にはヘッドマウントディスプレイや偏光メガネ、多数のプロジェクタや液晶ディスプレイを用いたものなど、様々なものの開発が進められている、また、多方向からの音情報提示に関しては映画館やホームシアターなどのように、一般的にも普及しているサラウンドシステムや、ヘッドホンを用いて任意の方向に音を知覚させることができる聴覚ディスプレイ技術などがある。このように、視覚や聴覚を対象としたバーチャルリアリティ技術は注目を集め、現在でも様々な研究が進められている。

視覚情報、聴覚情報に関するバーチャルリアリティ技術で共通して重要なのは、空間情報である。空間情報を把握する上で聴覚は重要な役割を果たしている。人間は振



り向かずに真横や真上，背後の物や景色を見ることはできないが，あらゆる方向の音を聞くことができる．さらに，真横や真上であっても到来する音の方向や距離を聞き分け，知覚することができる．つまり，人間が空間全体の情報を知覚する上では，聴覚こそ視覚を超える最も重要な器官である．したがって，バーチャルリアリティ技術のさらなる発展にとって，音空間情報提示技術の発展は重要である．

音空間情報を忠実に記録，提示するシステムとして音場再現という研究が行われ，注目を集めている．この技術は，実空間での音情報そのものを記録し，その音情報を人間の耳元で忠実に提示することによって実現できる．この技術によって，例えばコンサートホールなどの響きも含めたオーケストラ全体の音を別の場所，別の時間にて聞くことが可能となる．耳元での音が忠実に提示される原理に基づいたシステムであり，この音場再現は臨場感通信の中でも究極に位置する．

その上で本研究では，現状の音場再現技術にはない，さらに多様性に富んだ音空間提示システムを提案する．現状の音場再現は，空間全体の記録と再現に特化しているため，例えば記録したオーケストラの情報からバイオリンなどといった特定の楽器だけを取り出して提示することや，ホールの響きを取り去るといったことはできない．それに対し本研究では，空間全体の音を記録し，記録した音情報から各楽器の音と各楽器がホールを反射する音といった音の種類ごとに分解し，その中から任意の楽器音や反射音のみを提示できるようなシステムを対象として検討を進めることとした．このようなシステムを実現するためには「a：音空間全体を記録する技術」「b：記録した音空間から各楽器音や各反射音といった音の種類ごとに音を分解し，各音を抽出する技術」「c：抽出した各音情報を人間に対して忠実に提示する技術」の3つの要素技術が必要となる．このシステムが実現すれば，音空間情報からそれぞれ抽出した各音をすべて忠実に提示することにより音空間全体の音を再現できる他，たとえばオーケストラ演奏全体からバイオリン演奏だけを取り出して提示することや，そのバイオリン演奏を他のオーケストラ演奏と合わせて提示することも可能となる．このように任意の音だけを提示する技術や他イベントの音と組み合わせて提示する技術は，現在の音場再現では実現できない技術であり，より多様な臨場感を創出することが期待できる．

本研究では，このシステムの実現のための「a：音空間全体を記録する技術」と「b：記録した音空間から各楽器音や各反射音といった音の種類ごとに音を分解し，各音ごとを抽出する技術」の開発を行う．後者の「各音ごとを分離・抽出する技術」では，室内に話者や楽器といった音源が1つ存在する環境を想定し，その音空間で記録した室内の反射音を含む音情報から，音源から放射される音情報を抽出する技術の検討を行う．

話者や楽器といった実環境の音源は大きさや振動パターンを持つため，方向ごとに音の伝搬特性が異なる．普段の生活においても，話者の向きによって声の聞こえ方が異なることはだれもが経験していることである．これは，顔の正面方向では話者の口

から放射される音が直接伝搬するのに対して、顔の背面に行くにしたがって頭部を回折して伝搬する回折音の成分が多くなるため、音が小さくなり、波長が長く回折しやすい低音が強く聞こえるからである。人間は話者の近距離であれば話者の顔の向きも音の聞こえ方の違いから推定が可能であるという報告もある。また、楽器も方向ごとに異なった伝搬特性を持つ。この方向ごとの伝搬特性を放射指向特性と呼ぶ。先で述べた3次元映像提示などとの融合も考えた臨場感あるシステム構築のためには、記録した音情報から音源の放射指向特性を抽出し、提示することは重要である。放射指向特性の抽出や提示も各音源ごとに忠実に実現できれば、360°視点立体映像技術との組み合わせにより、自由視聴点システムの実現も夢ではない。また、CD音源の楽器の原音などに抽出した放射指向特性を付加することができれば、普段聞いているCD音源をコンサートホールに登場させることなども可能となる。将来これらの技術がすべて確立すれば、自宅などでも一流アーティストのライブやコンサートを臨場感を保ちつつ好きな視聴点で楽しむことができ、さらには複数アーティストのコラボレーションや、自宅にいながら他の演奏者の楽器音と合わせながらの練習なども実現できるようになるなど、バーチャルリアリティ音空間提示の夢は広がるばかりである。本研究はそのようなシステムを実現するための第一歩を踏み出す検討となる。

音源からの音を忠実に提示するためには「1: 音源の持つ放射指向特性」や、放射指向特性を持つ直前の音である「2: 原音情報」、音源が部屋のどこで鳴っているかという「3: 音源の位置情報」の3つが必要となる。本研究の目的は「a: 室内の音全体を記録すること」と、記録した音情報から1~3の属性を高精度に抽出することとするであり、これは「b: 記録した音空間から各楽器音や各反射音といった音の種類ごとに音を分解し、各音ごとを抽出する技術」のうち、音源に属する属性の抽出に当たる。

本論文の第1章では、本研究の研究背景と、実現したい音空間提示システムおよび必要関連技術、研究の目的、章ごとの具体的な検討内容についてを述べる。

第2章では、音空間全体を記録するためのシステムである包囲型マイクロホンアレイの構築とその仕様について述べる。以下の検討では、第2章において構築した包囲型マイクロホンアレイで記録した音情報を用いた検討を行う。

音源の放射指向特性を室内の反射音を含む音情報から抽出するためには、記録した音情報から音源の放射指向特性と室内の反射音特性を分離する方針で研究を行う。室内の反射特性は非常に複雑なため、記録した音情報のみから直接抽出することは困難である。よって、まず最初に音源の位置情報や原音情報を抽出し、それらの情報を用いて最後に放射指向特性の抽出を行う。

室内での音源の到来方向や位置の推定では、音源と波形が非常に似ている反射音の影響により精度が劣化することが知られている。第3章では、反射音に頑健な音源位置情報の高精度抽出に関する検討を行う。

残響除去に関する従来研究では，対象を音声信号としている場合が多いため，8～16 kHz をサンプリングレートとした方法が検討されてきた．しかし，本研究での最終目標は「音空間から好きな音を抽出する」ことであるため，抽出する音が音声であるとは限らず，物音や楽器音である場合においても精度よく残響を除去できる方法が求められる．よって，第4章では，残響除去の広帯域化について検討を行う．

第5章では，音源の放射指向特性の抽出法についての検討を行う．具体的には，抽出した原音情報と各観測点における観測信号から音源位置と各観測点間の伝達特性を抽出し，そこから放射指向特性の形成する伝達関数の抽出を試みる．

第6章では，本論文の結論を述べる．

# 目次

第1章 序論	1
1.1 はじめに	1
1.2 実環境の音空間を構成する属性	2
1.3 既存の音空間提示システムとその特徴	3
1.3.1 レンダリング型音空間提示システム	4
1.3.2 収音・再生型音空間提示システム：音場再現	5
1.4 次世代音空間提示システムの提案	6
1.5 本研究に必要な音空間の収音システム	9
1.6 室内音空間のモデル化	10
1.6.1 放射指向特性を考慮しない従来の多チャネル出力モデル	10
1.6.2 放射指向特性を考慮した多チャネル出力モデル	10
1.6.3 音源特性情報の高精度な抽出方法	12
1.6.4 放射指向特性の高精度抽出	14
1.6.5 原音情報の高精度抽出	14
1.6.6 音源位置情報の高精度抽出	15
1.7 抽出した属性値を用いた具体的な提示方法	16
1.8 本研究の目的	17
第2章 包囲型マイクロホンアレイの構築	19
2.1 はじめに	19
2.2 包囲型マイクロホンアレイ	19
2.3 157 ch 同時収音の同期確認	24
2.4 インパルス応答測定方法	27
2.5 まとめ	29
第3章 反射音に頑健な音源位置推定法の提案	31
3.1 はじめに	31
3.2 マイクロホンアレイを用いた音源位置推定	31
3.2.1 空間的なフーリエ変換に基づく方法	32

3.2.2	MUSIC 法	34
3.2.3	マイクロホンアレイを用いた音源位置推定	38
3.3	反射音による音源位置推定精度の劣化と空間平均化法	39
3.3.1	反射音による音源位置推定精度の劣化	39
3.3.2	空間平均化法の原理	42
3.3.3	不均一アレイへの応用	47
3.4	空間平均化法の音源位置推定への適用	48
3.4.1	音源位置推定への適用の問題点	49
3.4.2	3次元位置推定における空間平均化法の原理	49
3.5	RAP-MUSIC	53
3.5.1	仮想信号の平均化	53
3.5.2	RAP-MUSIC の提案	54
3.5.3	実験による提案法の性能評価	55
3.5.4	実環境における実験条件	56
3.5.5	実環境における実験結果	58
3.5.6	考察	67
3.6	まとめ	67
<b>第 4 章</b>	<b>室内残響下における原音情報の高精度抽出法の提案</b>	<b>69</b>
4.1	はじめに	69
4.2	MINT による完全な逆フィルタと残響除去	70
4.2.1	非最小位相系フィルタ	70
4.2.2	MINT	71
4.3	ブラインド多チャネルシステム同定によるインパルス応答の推定	73
4.3.1	ブラインド多チャネルシステム同定	73
4.3.2	LMS アルゴリズムに基づく方法	73
4.3.3	部分空間法を用いた伝達関数の推定	76
4.3.4	先行研究のまとめと問題点	79
4.4	ブラインド残響除去	80
4.4.1	Semi-blind MINT	80
4.4.2	線形予測を用いた高精度ブラインド残響除去	82
4.5	多チャネルブラインドシステム同定とブラインド残響除去の特徴	87
4.6	Pre-Whitening フィルタを用いた LIME アルゴリズムの高精度化	88
4.6.1	室内インパルス応答の測定	88
4.6.2	LIME アルゴリズムの問題点	89

4.6.3	White-LIME の提案 . . . . .	96
4.6.4	チャネル数の違いによる性能評価 . . . . .	101
4.6.5	高サンプリング周波数・高帯域信号を用いた場合の性能評価 . . . . .	109
4.6.6	マイクロホン配置の違いによる性能評価 . . . . .	109
4.7	まとめ . . . . .	115
<b>第 5 章</b>	<b>音源の放射指向特性の推定に関する検討</b>	<b>117</b>
5.1	はじめに . . . . .	117
5.2	放射指向特性の測定 . . . . .	118
5.2.1	無響室における放射指向特性の測定 . . . . .	118
5.2.2	室内における放射指向特性の測定 . . . . .	118
5.2.3	測定結果の比較 . . . . .	120
5.3	放射指向特性と室内残響分離の原理 . . . . .	122
5.4	原音情報を用いたインパルス応答の抽出 . . . . .	127
5.4.1	原音情報を用いたインパルス応答の抽出の原理 . . . . .	127
5.4.2	インパルス応答抽出の性能評価 . . . . .	127
5.5	インパルス応答からの放射指向特性抽出 . . . . .	129
5.5.1	インパルス応答切り出しによる放射指向特性の抽出 . . . . .	129
5.5.2	抽出した放射指向特性の性能評価 . . . . .	129
5.6	まとめ . . . . .	134
<b>第 6 章</b>	<b>結論</b>	<b>149</b>
<b>付 録 A</b>	<b>放射指向特性の時間波形，周波数特性および位相特性</b>	<b>155</b>
	<b>謝辞</b>	<b>169</b>
	<b>参考文献</b>	<b>171</b>



# 第1章 序論

## 1.1 はじめに

1876年，グラハムベルによる電話の発明によって，人々は空間を越えて会話ができるようになり，1877年には，トーマスエジソンによる蓄音機の発明によって，人々は時間と空間を越えて音を記録，蓄積，再生できるようになった．これらは音情報の記録，蓄積，伝送，再現の始まりである．

それから130年が経ち，2チャンネルのステレオを越え，現在では5.1 ch サラウンドシステムやそれ以上の多数のスピーカを用いたステレオフォニック的音響再生技術が主流である．また，反射のない自由空間における音源から聴取者の耳元までの音経路の伝達関数である頭部伝達関数 [1] を用いて音の位置を自由に変化させることが可能な聴覚ディスプレイという技術も登場し，ヘッドホンなどを用いたシステムも実現されている．

現在ではこれらの技術がホームシアターやヘッドホン製品として一般家庭にも普及しつつあり，音の動きや広がりを体験できるシステムが簡単に手に入るようになった．これは，人間が動きのある音や広がりのある音によって，あたかも本物の音が動いているような感覚や，その場にいるかのような音の広がりを音情報から知覚することで，あたかもその場にいるかのような高い臨場感が得られることも一因である．

ホームシアターやヘッドホン製品で提示される音空間は，基本的には制作者によって「加工された」ものである．この加工によって，映画館などでは人は実環境では体験できないような音空間を知覚することができる．

一方，まだ一般には普及していないが，音場再現という実環境の音空間を忠実に記録，提示する研究も古くから行われており，注目を集めている．この技術では，加工された音空間ではなく，実空間の音場そのものを提示できるため，高臨場感を与える提示システムであり，学術的な意義も非常に高い．

このように，現在の音響提示システムでは音の動きや距離，広がりといった音の空間性の再現と提示がこれまで以上に重要視されている．

本研究では，より高詳細な音空間提示のために，実環境の音空間全体を記録し，そこから音空間を構成する属性を属性ごとに分離し，その属性の中から必要な属性のみを編集，加工し，提示するシステムの実現を目指す．このシステムが実現すれば，人



は実空間のある特定の音のみを聞くことができる，さらには，様々な空間での音を組み合わせ提示するなどといった実環境の音空間を超えた提示なども可能となり，既存のシステムを超える高臨場感音空間提示が期待できる．

このシステムの実現のためには，属性を抽出する技術とそれらの知覚的提示技術が必要になる．本研究では，音空間全体の記録と，記録された情報からの属性の分離，および属性値の抽出を行う技術の開発を目的とする．

以降 1.2 節では，実環境の音空間や音源自体の属性を高精度に抽出するために，まずは実環境の音空間を構成している属性について述べる．1.3 節では，現状での実環境音空間提示システムとその特徴について述べる．1.4 節では，1.3 節を受けて，本研究で目指す今までにはない新しい音空間提示システムとその特徴について述べる．1.5 節では，新しい音空間提示システム実現のための根幹である，音空間を構成する属性を分解，抽出する技術について述べ，本研究での具体的方針を明らかにする．最後に，1.6 節では，本研究の目的および章構成について述べる．

## 1.2 実環境の音空間を構成する属性

実環境の音空間の特徴を綿密に把握するために，まずは実環境における音空間を構成する属性について整理する．音空間 (室内) を構成する要素は以下の 5 つであると考える．

1. 音源の位置
2. 音源が発する原音
3. 放射指向特性
4. 初期反射音
5. 後部残響音

この 5 つのうち，1. から 3. は音を発生する物体，すなわち音源が持つ特性であり，4. と 5. は空間 (=室内) に備わる特性であると大まかに言える．

まず，1. は源の位置情報であり，部屋のどこで音が鳴っているかを示す．2. は，物体の振動から放射される瞬間の音 (=原音) であり，どんな音が鳴っているかを示す．3. は，音の向きによる放射される音のレベルの違いである．これは実環境の音が，音源自体に大きさがあることや，振動パターンが点音源とは異なるため，音源と距離が等しい位置であっても，方向によって音の伝搬特性が異なることに起因する．4. は，音源から放射された音が壁面に反射して伝搬される音である．そして 5. は，反射や拡散など

を繰り返し，方向性がなくなった音であり，後部残響と呼ぶ．例えば，残響時間とは室の響きの特性を表す 1 つの指標である [2] ．

図 1.1 に音源をスピーカとした場合における室内音空間を構成する 5 つの属性を説明する図を示す．室内でのスピーカの配置位置が 1. 音源位置情報，スピーカから放射された直前の音が 2. 原音情報である．スピーカの方角によって異なる伝搬特性が 3. 放射指向特性を表す，スピーカから放射された音が壁面を反射し，4. 反射音が生成され，反射音が反射や拡散などを繰り返し，5. 後部残響となる．

室内音場は主としてこれら 5 つの要素から構成されており，このうちの 1 つでも変化すれば，音空間は別のものとなる．実環境は，この 5 つの特性が音源の個数分だけ混合された形で存在し，極めて複雑なものである．

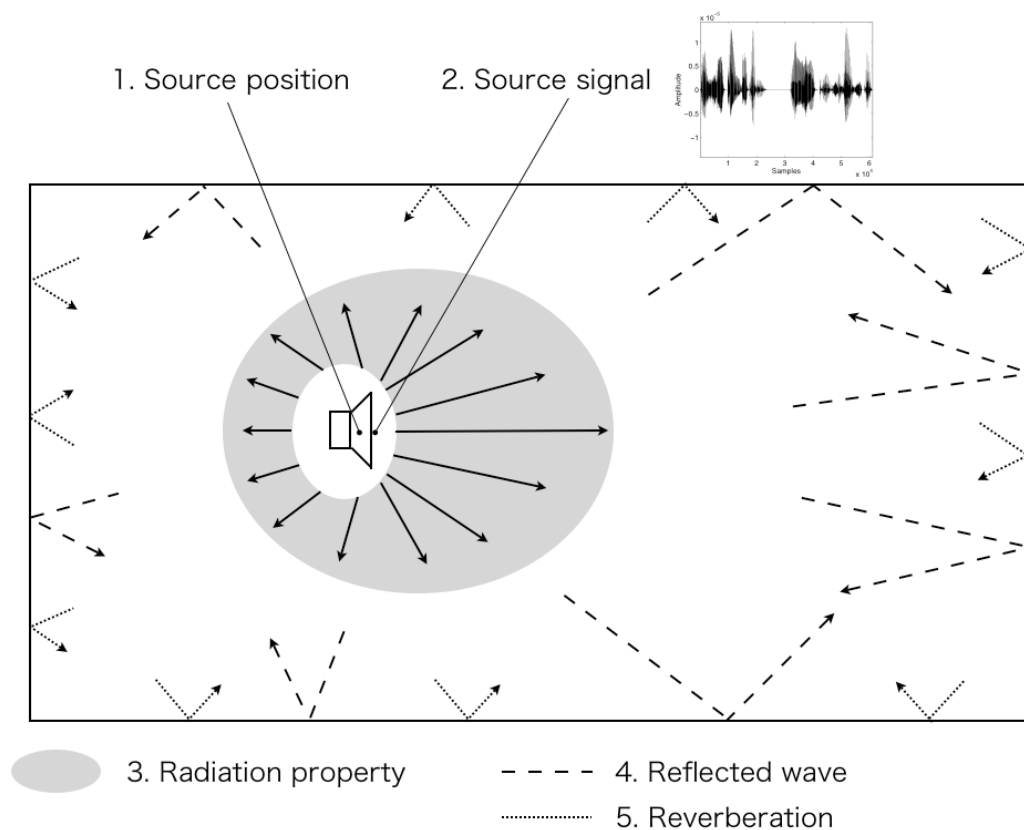


図 1.1: 音空間を構成する 5 つの属性

### 1.3 既存の音空間提示システムとその特徴

人間は，耳に入る音源からの直接音の情報から，音源までの距離や方向，反射音の情報から部屋の大きさといった空間情報を認識することができる．これは，人間が両

耳に届く音の時間差やレベル差の分析に基づいて、音源の距離や方向を知覚することができるためである。この能力を音像定位と呼ぶ [3]。

このしくみを利用し、人間の両耳に届く音の時間差や音圧差を積極的に制御することで、人間に対して音の空間情報を知覚させることができる。

このシステムは、大きくわけて2つに類型することができる。1つは、原音に加工を加えて3次元的な音空間を提示するシステムであり、もう1つは、実環境の空間情報を厳密に記録、提示するシステムである。以下ではこの2つのシステムの特徴について述べる。

### 1.3.1 レンダリング型音空間提示システム

本研究では、原音に加工を加えて3次元的な音空間を提示するシステムを総称して、レンダリング型音空間提示システムと呼ぶこととする。

1.1 節でも取り上げた 5.1 ch サラウンドシステムや聴覚ディスプレイ技術はこのシステムに属する。5.1 ch サラウンドシステムなどはスピーカの数少なく、音の動きなどを擬似的に再現する方法であるが、聴取者の周囲にさらに多数のスピーカを配置し、原音場で到来する方向の音をその方向に対応するスピーカから出力することにより、精度よく音空間を提示することが可能となる。この技術を立体角分割法といい、多くの研究がなされている [1, 4, 5]。

また、頭部伝達関数に基づいて、人間の耳元の音圧をヘッドホンなどを用いて制御することにより、任意の位置に仮想音像を提示することができる [6]。さらに、ヘッドホンを用いた聴覚ディスプレイに位置センサを導入し、聴取者の動きにも対応した頭部運動感応型聴覚ディスプレイも開発されている [5, 7–13]。この技術を用いることにより、人間の頭部運動にも対応した任意の位置に仮想音像を提示することが可能となる。

ヘッドホンを用いた聴覚ディスプレイ技術は伝達関数に基づく方法のため、伝達関数の記録と再現を忠実に行うことができれば、人間に対して任意の位置に、物理的に厳密な仮想音像を提示することが可能になる。このため、3次元音空間提示の基盤となる技術と言える。またこの技術は、頭部伝達関数に畳み込む音源を、無響室で収音した実環境の原音や、コンピューターシミュレーションなどで作成した反射音などすることで、どんな音でも仮想音像として3次元提示することが可能である。さらに、文献 [12] で提案されている聴覚ディスプレイは、虚像法 [14] を用いた反射音や、音源や聴取者の動きによるドップラー効果を模擬することができる。

### 1.3.2 収音・再生型音空間提示システム：音場再現

レンダリング型音空間提示が任意の位置の音源を提示するシステムであるのに対して、実環境の音空間全体を厳密に提示することのできる収音再生型音空間提示システムに関する研究も行われている。この技術を音場再現と呼ぶ。この技術の基本原理は、人間の耳元の音場や、さらには人間を取り囲む空間 A 全体の音場を、別の実空間 B の音場とまったく同じものとして提示することである。人間の耳に入力される音情報が再現される別空間 B のものと物理的にまったく同じであれば、人間はその音を空間 A で聞いたとしても、あたかも別空間 B にいるかのような知覚をすることになる。音場再現に関する従来方法には、大きく分けて 2 つある。

その 1 つに、1933 年にベル研究所の Fletcher らが始めたダミーヘッド (図 1.2) によるバイノーラル録音がある。ダミーヘッドの耳元で録音した、2 チャンネルの信号をヘッドホン再生すること (バイノーラル再生 [15, 16])、または多数のスピーカを用いて聴取者の耳元の音圧を制御すること (トランスオーラス再生 [17]) により、聴取者はあたかもダミーヘッドを設置した空間にいるかのような音空間を体験することができる。

この技術は容易に実現可能ではあるが、ダミーヘッドの耳元 2 点の音情報しか再現できないため、聴取者の動きには対応できないという課題がある。この課題に対して戸嶋らは、聴取者の頭部運動にも対応したテレヘッドを開発している [18]。

バイノーラル録音のような点制御の方法に対して、ある閉領域内の音空間全体を再現する方法として、波面合成法 [19–21] や Kirchhoff-Helmholtz 積分方程式に基づく境界音場制御 [23, 24] による再生方法がある。

式 (1.1) に Kirchhoff-Helmholtz 積分方程式を示す。

$$p(s) = \int \int_S G(\mathbf{r}|\mathbf{s}) \frac{\partial p(\mathbf{r})}{\partial n} - p(\mathbf{r}) \frac{\partial G(\mathbf{r}|\mathbf{s})}{\partial n} \delta S \quad (1.1)$$

ここで  $G(\mathbf{r}|\mathbf{s})$  は 3 次元空間中の点音源のグリーン関数であり、3 次元空間では、

$$G(\mathbf{r}|\mathbf{s}) = \frac{1}{4\pi|\mathbf{r} - \mathbf{s}|} e^{-j\omega \frac{|\mathbf{r} - \mathbf{s}|}{c}} \quad (1.2)$$

と表される。ここで、 $p$  は複素音圧、 $\mathbf{r}$  は振動表面  $S$  上の点、 $\mathbf{s}$  は音場の点、 $\partial/\partial n$  は  $S$  からの外向き法線に関する微分を表す。式 (1.2) のグリーン関数は 3 次元自由音場内において点  $\mathbf{r}$  に点音源がある場合の  $\mathbf{s}$  における音圧を示す式に等しい。Kirchhoff-Helmholtz 積分方程式による境界音場制御はこの方程式に基づき、境界上の音圧  $p$  と音圧勾配  $\partial p(\mathbf{r})/\partial n$  を制御することにより、領域  $S$  内の音場を再現する (図 1.3)。実際には、図 1.4 に示すように近接する観測点間の音圧差を音圧力勾配とする。つまり、観測点で観測された音圧を別の場所で忠実に再現することにより、領域  $S$  内の音場を再現する。そのためには、あらかじめスピーカ  $i$  と制御点  $j$  の伝達関数  $g_{ij}$  をすべて測定し

ておく必要がある (図 1.5) . この方法により , 領域内の音空間を忠実に再現することができる . 近年では , この境界音場制御によって , 離れた 2 つの空間同士において , 第 3 の別の音空間を共有するといったシステムも提案されている [25, 26] . ただし , これらの方法の課題は , 音源が制御領域内部にある場合は式 (1.1) が一意に解けなくなるため , 記録する領域内に音源を含むことができないということがあり , 記録領域が限定される点にある .



図 1.2: ダミーヘッド (高研製 SAMRAI)

音場再現の最大の特徴は , 1.2 節で示したように実環境での複雑な音空間を , 制御点の記録と提示という作業のみで音空間全体の再現が可能である点である . つまり , 記録した音情報自体には逆フィルタ処理以外一切の処理を施す必要がない .

しかし逆に , 現状の音場再現は 1.2 節で示した 5 つの属性を分解することや , 分解した属性を別々に提示することなどはできない . この点において音場再現は , 空間全体を再現することに特化したシステムであると言える .

## 1.4 次世代音空間提示システムの提案

1.3 節で述べたように , 現状の音空間提示システムでは , 音を加工して任意の位置に提示することに特化したレンダリング型と , 音空間全体の提示に特化した音場再現型とが存在し , それぞれ人間に対して高い臨場感を与えることができる . これらの技術は , さらなる高臨場感通信実現のため , 現在でも様々な研究が行われている .

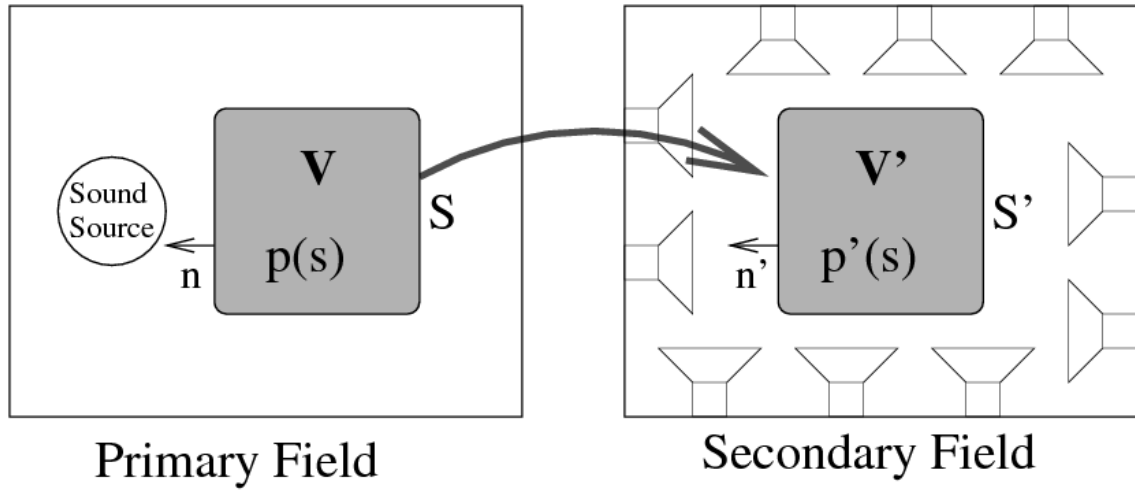


図 1.3: 境界音場制御による音場再現の概念図 [23]

$p(s)$  : 閉空間  $V$  , 閉領域  $S$  内の音圧 ,    $p'(s)$  : 閉空間  $V'$  , 閉領域  $S'$  内の音圧  
 $n$  : 領域  $S$  の法線ベクトル ,    $n'$  : 領域  $S'$  の法線ベクトル

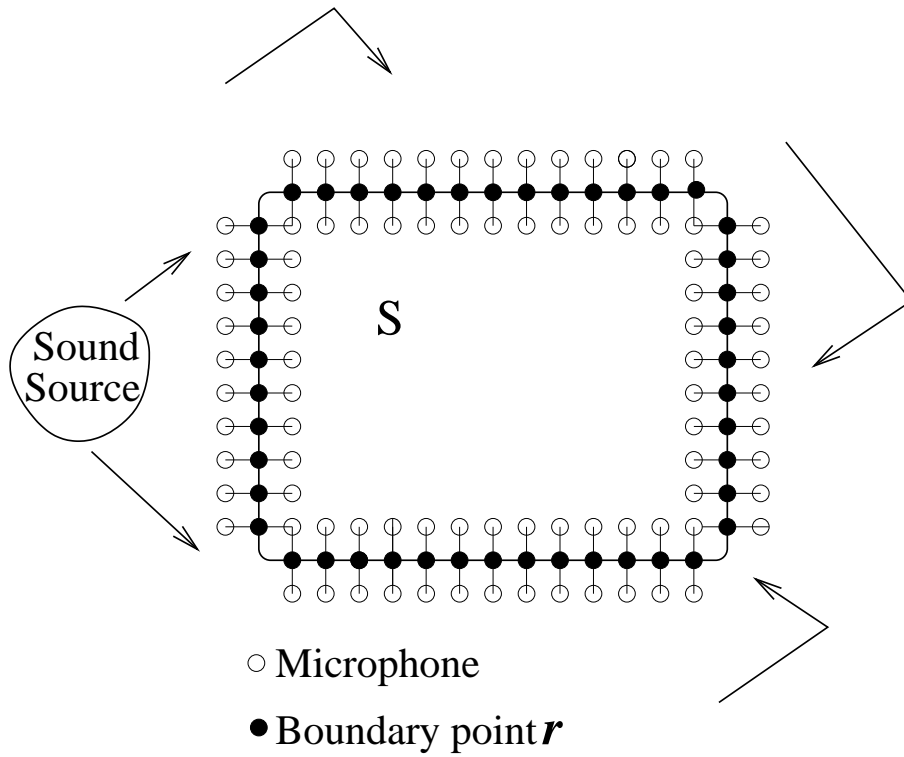


図 1.4: 閉領域  $S$  境界面の音圧および音圧力勾配の計測 [23]

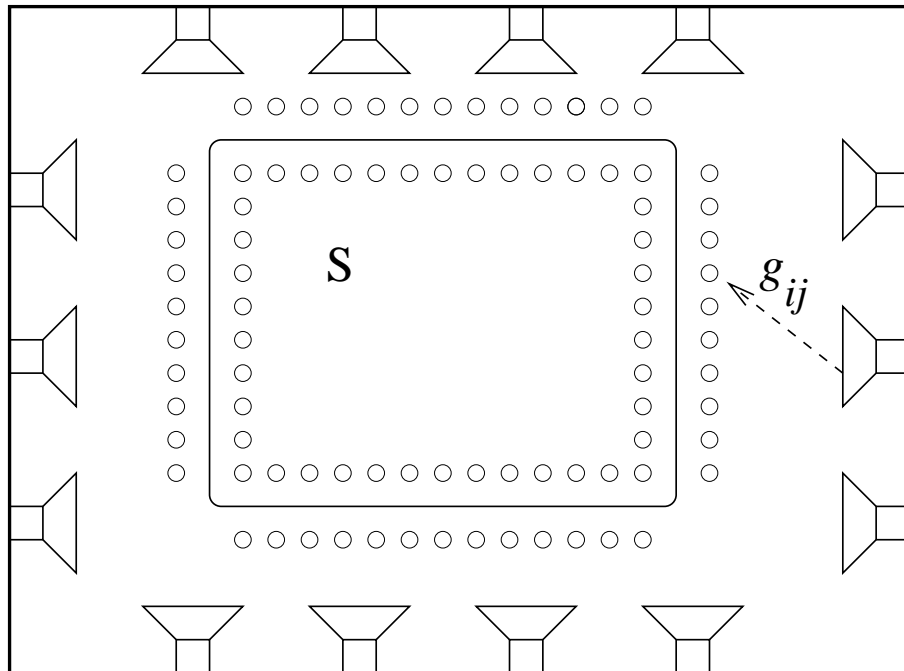


図 1.5: 制御点とスピーカ間の逆フィルタ  $g_{ij}$  を用いた逆システムによる閉領域  $S$  内の音場再現 [23]

それに対して本研究では，さらに多様性に富んだ臨場感通信実現のために，従来のシステムにはない，まったく新しい音空間提示システムを目指す．本研究で目指す音空間提示システムとは，音空間全体を記録し，記録した音空間を構成する音源位置情報，原音情報，放射指向特性，反射音，後部残響の 5 つの属性を属性ごとに分解し，必要な属性を抽出，編集，提示するものである．このシステムが実現すれば，従来方法では実現できなかった以下のような音空間を創出することができる．

- i) 実環境の任意の音源の音のみを忠実に提示
- ii) 他イベントの音環境の音源の音と同時に提示
- iii) 他イベントの音環境の音源の属性を入れ替えての提示
- iv) 仮想的な属性を実環境の音源に組みこんだ提示

この他にも多様な提示が可能となり，これらを実現することができると，音空間を構成する属性の制御が可能となり，例えば以下のような音空間を創出することができるようになる．

- i-ex) 収録したオーケストラの中からバイオリンやトランペットといった特定の楽器のみの音を取り出し，提示

- ii-ex) 別のオーケストラ同士の共演といった実環境では存在し得ない音空間の実現
- iii-ex) さらに別の会場の音空間特性にした提示
- iv-ex) コンピュータで作成した音源や CD など別の音源に実環境の音空間を適用した，実環境と仮想環境の音空間の融合

このような音空間を実現できれば，空間内から必要な音だけを聞くことや，さらに他の空間の音との組み合わせなども可能となり，今までにはない多様性に富んだ臨場感を得ることが期待できる．

そのためには，「(a) 音空間全体を記録する技術」，「(b) 記録した音情報から音空間を構成する 5 つの属性を分解，抽出する技術」，「(c) 抽出した属性を属性ごとにレンダリングする技術」が必要となる．そこで本研究では，(a) と (b) の開発を行う．

その上で，本研究ではまず最初に，i) を忠実に実現するための属性の抽出を目標とする．記録した音空間全体の情報には 5 つの属性が含まれているため，まずは音源の属性と空間の属性を分離する必要がある．その上で重要であるのが，放射指向特性の抽出となる．実環境に存在する音源は点音源ではないため，必ず方向ごとに放射指向特性を持つ．また，文献 [28–30] では，人間は話者の向きを音声から判断することができるという報告もなされている．よって，空間情報を厳密に再現するためには，放射指向特性の抽出および提示は重要であると言える．放射指向特性をも厳密に再現することができれば，3 次元立体映像提示システムと組み合わせることにより，任意の位置から仮想空間の視聴を可能とする自由視聴点システム [31] への応用など，さらなる臨場感通信が期待できる．

よって，上記のような新しい音空間提示システム実現のためには，音空間全体の記録と，記録した音空間からの音源が持つ属性 (位置情報，原音情報，放射指向特性) の抽出がまず必要であると考えられる．

## 1.5 本研究に必要な音空間の収音システム

音空間を構成する属性を高精度に抽出するためには，音空間を記録する技術と，記録された音空間情報から属性を分離，抽出する技術が必要となる．本節では，音空間情報の記録と，記録された音空間情報から音源の持つ属性のみを分離，抽出する指針と解決すべき問題点について述べる．

音空間を構成する属性を高精度に抽出するためには，音空間の情報を余すところなく記録することが求められる．従来法である音場再現は，ダミーヘッドなどでは 2 チャネルであり，境界音場制御などでは制御領域を囲む境界上の音圧変化を記録するシス



テムであった．しかし，これらの記録形態では，観測点が音源を囲んでいないため，方向ごとの放射指向特性を含む音源の情報を余すところなく記録することはできなかった．そこで，本研究では積極的に音源を囲んだ観測点配置とし，音源の情報を余すところなく取り込む形態とする．音源を囲む形で收音することにより，音源から放射される様々な方向の音を取り込むことが可能となるため，様々な方向の音源の放射指向特性を記録することができる．この記録された情報から，放射指向特性を取り出すことができる．これらの観測点で記録された音空間情報から音源の持つ属性の分離と抽出を行う必要がある．

## 1.6 室内音空間のモデル化

本研究の目標は，図 1.1 に示したような直接音と反射音の存在する室内の観測信号のみから音源特性を属性値として抽出する問題に帰着できる．

室内の反射音や残響が存在する中で音源情報を抽出する従来研究はこれまでにたくさんのものである．次項以降，それらを概観する．

### 1.6.1 放射指向特性を考慮しない従来の多チャネル出力モデル

観測点にて記録された観測信号のみから音空間構成要素の分離と抽出を検討するにあたって，システムをモデル化する．室内における観測点への出力システムは，音源(=入力信号)が一つの場合は，図 1.6 のような Single-Input Multi-Output(SIMO) モデルとして記述することができる． $s(n)$  が原音信号であり， $x_i(n)$   $i = 1, 2, \dots, N$  は各マイクロホンにおける観測信号(チャネル数  $N$ )， $b_i(n)$   $i = 1, 2, \dots, N$  は各マイクロホンに入力される背景雑音，そして  $H_i(n)$   $i = 1, 2, \dots, N$  が室内伝達関数を表す．室内伝達関数とは，室内の反射や残響成分であり，図 1.7 のようなインパルス応答  $h(n)$  で記述される．

多チャネルの観測信号からインパルス応答や音源，位置情報を抽出するといった先行研究は，これまでも到来方向推定，残響除去，音源分離，システム同定など，非常にたくさんのものである [32, 33]．しかし，高精細な音源提示には放射指向特性の抽出は必要不可欠であり，入出力のモデル内に組み込む必要がある．

### 1.6.2 放射指向特性を考慮した多チャネル出力モデル

本研究は，音源に放射指向特性があることを仮定し，その特性を抽出することが目的とする．また，観測点は音源を囲んで配置することを想定しているため，音源の前

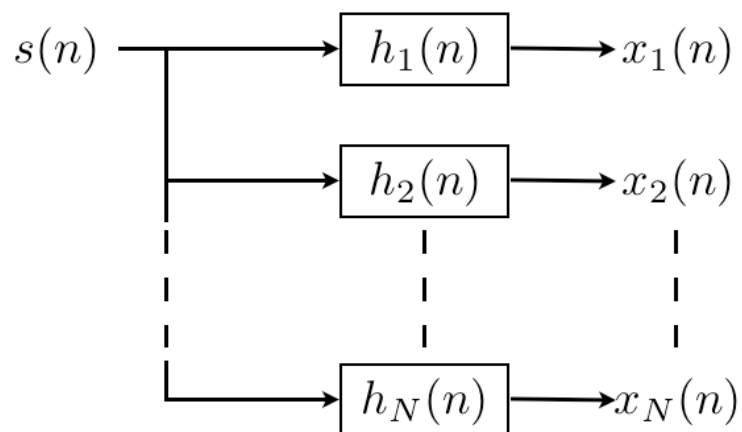


図 1.6: SIMO (Single-Input Multiple-Output) モデル  $s(n)$ : 原音信号,  $h_i(n)$ : 音源位置から観測点  $i$  までの室内伝達関数,  $x_i(n)$ : 観測点  $i$  での観測信号

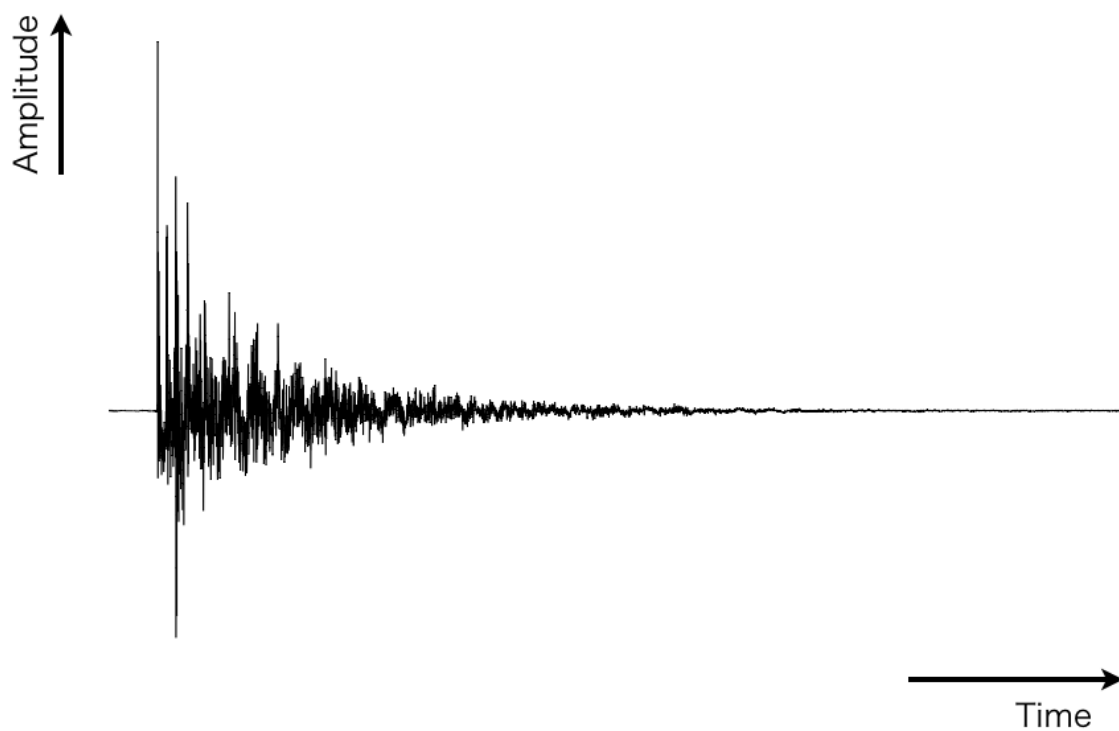


図 1.7: 室内インパルス応答  $h(n)$  の例

方だけではなく，音源の横方向，後方から届く音も記録する．よって，各観測点には，音源からの放射指向特性も入力される．しかし，音源の原音自体は共通の  $s(n)$  であるため，各方向の放射指向特性は別の伝達関数として考えることができる．よって，従来のマイクロホンアレイシステムとは異なり，本研究では，図 1.8，1.9 のように各方向の放射指向特性の伝達関数 (= インパルス応答)  $h_D$  と室内インパルス応答  $h_R$  との混合モデルであると仮定し，放射指向特性  $h_D$  を別途抽出する問題と考える．

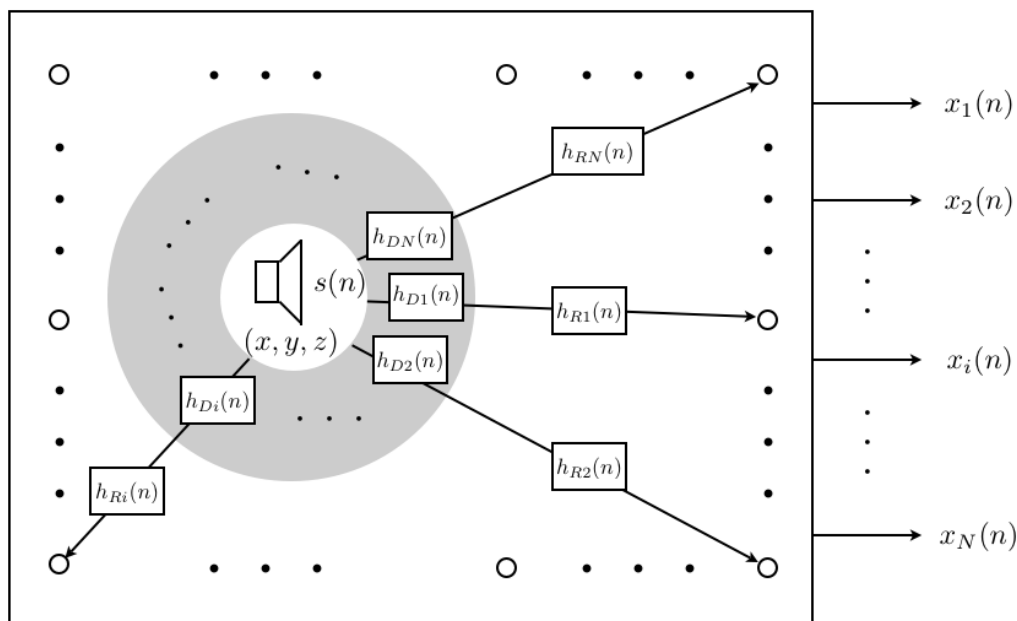


図 1.8: 放射指向特性を考慮した室内 SIMO モデルの概念図  $s(n)$ : 原音信号,  $h_{D_i}(n)$ : 放射指向特性の伝達関数,  $h_{R_i}(n)$ : 室内インパルス応答,  $\circ$ : 観測点,  $x_i(n)$ : 観測信号

### 1.6.3 音源特性情報の高精度な抽出方法

観測信号から音源特性の属性値を抽出することは，放射指向特性を含んだ図 1.8 の SIMO モデルから，1. 音源位置情報 (3 次元座標情報)，2. 原音情報 (波形情報)，そして 3. 放射指向特性 (インパルス応答) を属性値として抽出することである．

本項では，観測信号からこれらの属性値を抽出する具体的な方針について述べる．

マイクロホンを異なる位置に複数配置し，観測された信号の位置情報を用いて信号処理を行い，音の到来方向を推定することをマイクロホンアレイ信号処理と呼ぶ．マイクロホンアレイ信号処理によって，音源の位置情報や原音抽出が可能となるが，精度がマイクロホンの配置する間隔に依存し，間隔が大きい場合，空間周波数の折り返し歪みのため，抽出精度が劣化することが知られている．

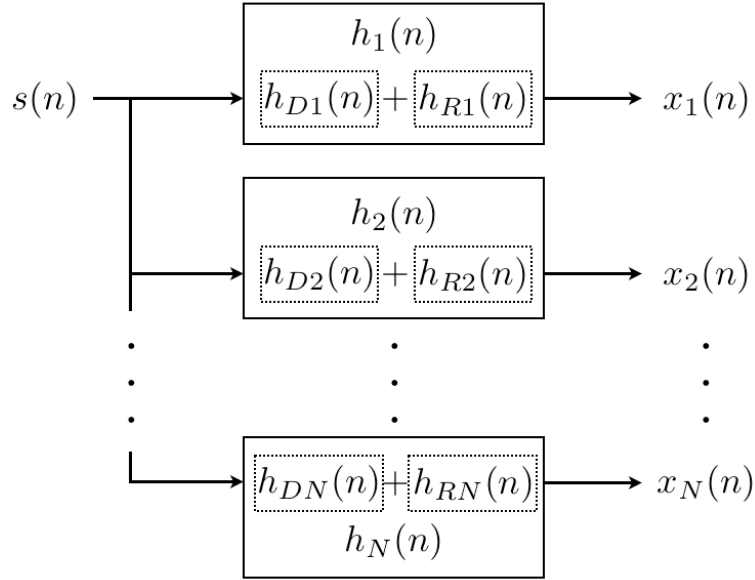


図 1.9: 放射指向特性を考慮した SIMO モデル  $s(n)$ : 原音信号,  $h_{Di}(n)$ : 放射指向特性の伝達関数,  $h_{Ri}(n)$ : 室内インパルス応答,  $x_i(n)$ : 観測信号

これに対して、マイクロホンの位置情報を利用せずに、室内インパルスを観測信号のみから直接求めるブラインドシステム同定 [34] に基づく方法がある。ブラインドシステム同定とは、観測信号  $x_i$  の情報のみを用いて系のインパルス応答  $h_i(n)$  を推定する方法であり、マイクロホンの位置情報は必要としないため、ブラインド信号処理に属する。ブラインドシステム同定は抽出精度はマイクロホン間隔に依存しないため、音源位置情報や原音情報の高精度な抽出が可能となる。

実環境での適用を考えた場合、マイクロホン間隔を細かく取ることはあまり現実的ではない。また、より高精度な属性値を抽出するためにも、マイクロホン間隔に依存しない方法が望まれる。よって、本研究での原音情報、そしてインパルス応答の抽出については、マイクロホン間隔に依存しないブラインド信号処理の原理に基づく方法を用いた検討を行う。

その上で、はじめに観測信号  $x_i$  のみからブラインドシステム同定によってインパルス応答  $h_i(n)$  を抽出する策略を取ることが、本研究において重要である。まず、放射指向特性を属性値  $h_{Di}(n)$  として抽出するためには、系全体のインパルス応答  $h_i(n)$  から放射指向特性の属性値  $h_{Di}(n)$  を分離する必要があるため、 $h_i(n)$  を抽出することは必須となる。また、音源と各観測点間の  $h_i(n)$  を抽出することができれば、 $h_i(n)$  の逆フィルタを算出することにより、原音情報の属性値  $s(n)$  を高精度に抽出することが可能となる。また、各  $h_i(n)$  の直接音の到来時間差情報から、音源位置情報の属性値  $(x, y, z)$  を抽出可能である。

しかし，音響分野でのブラインドシステム同定は，以下の2点において非常に困難であるとされている．一つは，系のインパルス応答は残響時間で規定されるため，応答長が非常に長いという点である．例えば，残響時間が0.1 sと短いものであっても，サンプリング周波数が高くなるにつれて数千タップと非常に大きなタップ数となる．もう一つは，その大きな応答の正確な応答長が未知である点である．詳しくは第4章にて述べる．

これに対して，観測信号  $x_i$  のみから系のインパルス応答  $h_i(n)$  の逆フィルタを直接推定し，原音情報  $s(n)$  を直接推定するブラインド残響除去という方法がある．この方法では系のインパルス応答長が非常に長い点や応答長が未知である点に対して頑健である．ただし，ブラインド残響除去では，音源に一番近い観測点がどれであるかといった先見的な情報を必要とする．

以上の理由から，本研究では，音源位置情報  $(x, y, z)$  を最初に抽出する．位置情報と観測点の位置情報から音源に一番近い観測点の情報を算出し，ブラインド残響除去によって原音情報  $s(n)$  を抽出する．そして最後にこれらの情報を用いて，放射指向特性  $h_{Di}(n)$  の抽出を行う．以上のような順番で信号処理を進めていく．

これらを抽出する上で具体的な解決すべき点を次項で整理する．

#### 1.6.4 放射指向特性の高精度抽出

まず，室内における音源位置情報抽出で求められることは，反射音に対する頑健性である．反射音は原音と非常に相関が高いため，反射音の存在する環境での位置情報抽出は精度が劣化することが知られている．マイクロホンアレイの従来研究の多くは，音源の到来方向推定，つまり音がどの方向から到来したのかを推定する検討が主流である．反射音に頑健な推定方法に関する検討も，基本的には到来方向推定にしか行われておらず，3次元位置推定での検討はなされていない．しかし，本研究で抽出すべき属性値は，反射音が存在する中での音源の3次元位置情報  $(x, y, z)$  である．よって，音源位置情報抽出において具体的に解決する点は，反射音に頑健な3次元位置情報抽出方法の開発である．

#### 1.6.5 原音情報の高精度抽出

室内における原音情報抽出で求められることは，ブラインド残響除去の広帯域化である．系のインパルス応答長が非常に長い点や応答長が未知である点に対して頑健であるブラインド残響除去は，音声認識などを後段に想定した音声信号の残響除去を目的として行われているため，使用されているサンプリング周波数は8 kHz～16 kHzで

ある．しかし，本研究で求められる原音情報は，音声だけでなく，室内で発生する音一般であり，肉声感や場の臨場感を伝えるためには，さらに高いサンプリング周波数での残響除去が必要である．よって，原音情報抽出においては，ブラインド残響除去の広帯域化，高精度化が具体的な課題である．

### 1.6.6 音源位置情報の高精度抽出

放射指向特性の抽出は基本的にほとんど検討されていない問題である．本研究では，抽出した原音  $s(n)$  と観測信号  $x_i$  からインパルス応答  $h_i(n)$  を求め，その応答から放射指向特性  $h_{Di}(n)$  のみの抽出を試みる．その際，抽出したインパルス応答  $h_i(n)$  は音源と観測点までの距離  $r_i$  によって距離減衰量が異なるため，距離減衰を補正する必要がある，よって，放射指向特性を抽出するためには音源の位置情報も必要となる．

以上のようにして観測信号から 3 つの属性の高精度な抽出を行う．具体的な抽出の流れを図 1.10 に示す．

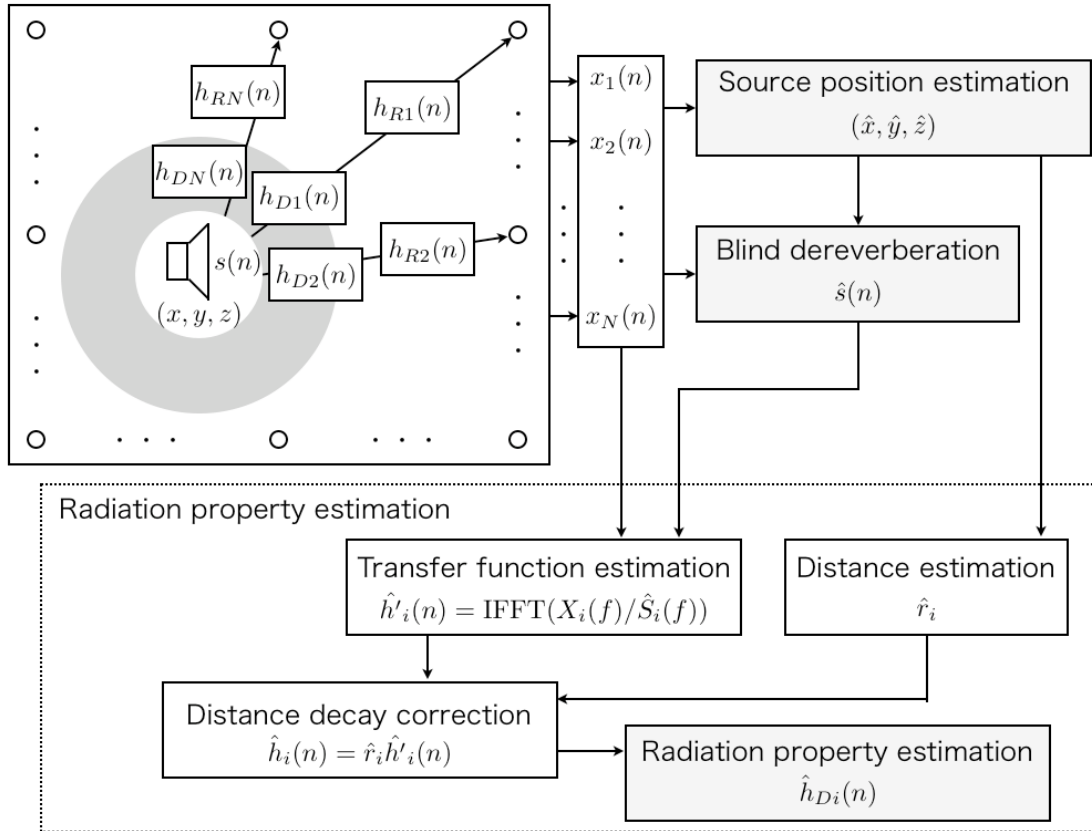


図 1.10: 観測信号  $x_i(n)$  から各属性値 (音源位置情報，原音情報，放射指向特性) 抽出のブロック図 ( $x, y, z$ ): 音源位置座標， $s(n)$ : 原音情報， $h_{Di}(n)$ : 放射指向特性， $h_{Ri}$ : 室内伝達特性

## 1.7 抽出した属性値を用いた具体的な提示方法

図 1.10 に示した各属性値  $(x, y, z)$ ,  $s(n)$ , そして  $h_{Di}(n)$  を抽出することができれば, これらの属性値を用いて室内の音源のみを高精度に提示することが可能となる. 既存の聴覚ディスプレイや放射指向特性を再生できるスピーカ [35] などを用いて提示が可能となる. ここでは, 聴覚ディスプレイを用いて提示する例を図 1.11 に示す. 従来の聴覚ディスプレイは点音源提示であるが, 放射指向特性がインパルス応答  $h_{Di}(n)$  として抽出されるため, 聴取者に対して放射指向特性を含んだ提示が可能となる. さらに, このシステムに聴取者の位置情報を取得する位置センサを導入することにより, 聴取者の位置の変化に追従して放射指向特性や音圧レベルを変化させることができる. つまり, 聴取者が提示される音源に近づくことや, 周囲をまわることが可能となる.

さらに, 1.4 節で示した通り, 音源特性を属性値  $(x, y, z)$ ,  $s(n)$ ,  $h_{Di}(n)$  として抽出しているため, 他のイベントの属性値と入れ替えることや, 属性値自体の編集も可能となる. これによって, さらなる高臨場感音空間提示が期待できる.

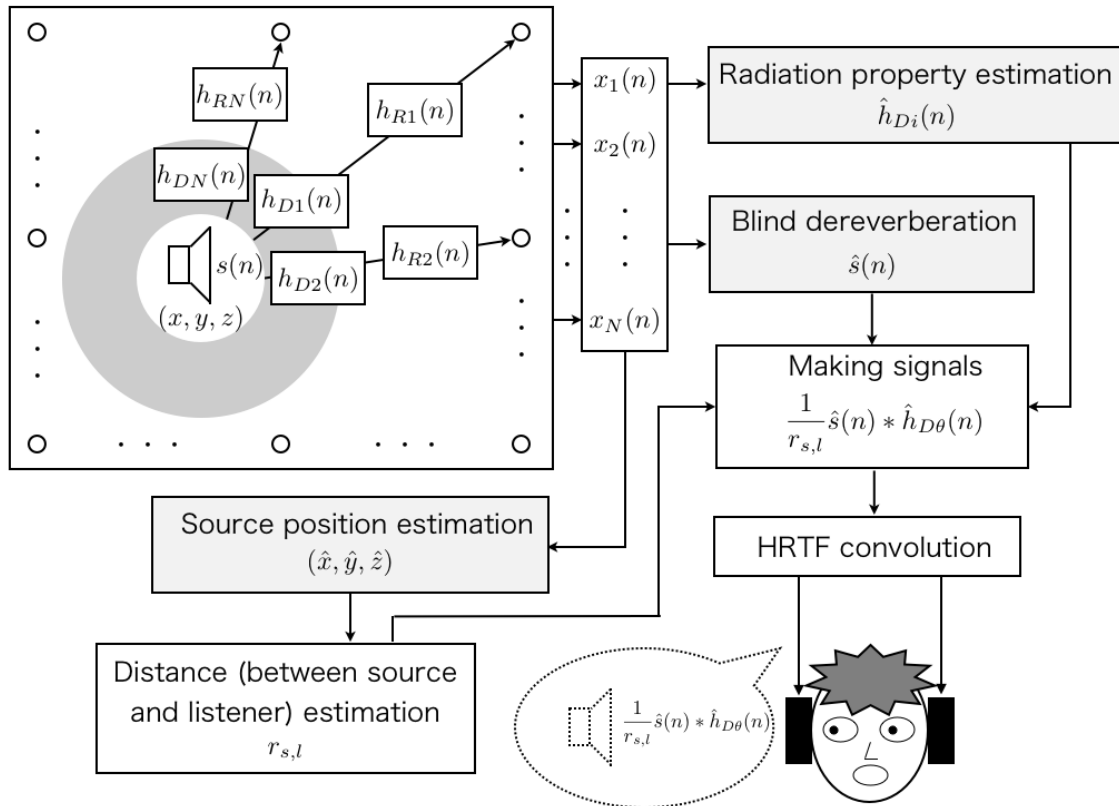


図 1.11: 観測信号のみから抽出した属性値を用いた聴覚ディスプレイによる音空間提示の概観

## 1.8 本研究の目的

音源の位置や動きなどの空間情報を含んだ音空間情報の記録や提示が近年注目を集めている．従来の音空間提示システムには，音空間を加工して仮想的に提示するレンダリング型システムと音空間全体を記録，再現する音場再現システムが存在する．既存の音場再現が音空間全体を提示するシステムであるのに対して，本研究では，記録した音空間から音空間を構成する音源位置情報，原音情報，放射指向特性，反射音，後部残響の 5 つの属性を属性ごとに分解し，必要な属性のみを抽出，編集，提示できるようなシステムを目指す．このようなシステムが実現すれば，実空間の任意の音だけを忠実に提示することや，他の音空間の音情報と組み合わせて提示することや，属性値を加工して提示することが可能となり，今までにはなかった多様な臨場感を創出できると期待できる．

新しい音空間提示システム実現のためには，音空間全体を記録する技術，記録した音情報から属性値を抽出する技術，そして属性ごとの忠実なレンダリング技術の 3 つが必要となる．そこで本研究では，室内音空間全体の記録と，記録された室内音情報から音空間を構成する 5 つの属性値のうち，音源の持つ属性である音源位置情報，原音情報，そして放射指向特性を属性値として高精度に抽出することを目的とする．その中でも放射指向特性は音源の向きを属性であり，厳密な音空間の提示のためには重要な属性値である．

観測信号のみから放射指向特性の属性値を抽出するためには，放射指向特性の形成する伝達関数と室内伝達関数の形成する伝達関数が混合された伝達特性を抽出し，その特性から放射指向特性の伝達関数のみを抽出する必要がある．しかし，室内伝達関数は応答が非常に長いため，観測信号のみの情報から直接抽出することは困難である．そこで，まずは音源の位置情報，原音情報を抽出し，これら 2 つの情報を用いて放射指向特性の抽出を行う．

以下，論文構成について述べる．

第 2 章では，音空間全体を記録するための，室内全体をマイクロホンで囲んだ包囲型マイクロホンアレイの構築について述べる．包囲型マイクロホンアレイは計 157 ch の多チャンネルで構成されるため，全チャンネルを時間的に同期させる必要がある．そこで，157 ch を用いたマイクロホンアレイのシステム構成，同期の確立方法および同期の確認方法について説明する．以降は，この設備を実験機材として，具体的な属性の抽出方法を検討する．

第 3 章では，室内における音源位置情報の高精度な抽出法についての検討を行う．室内での音源位置推定では，直接音と相関の高い反射音の存在により推定精度が劣化する問題がある．反射音に頑健な従来法も提案されているが，それは音源の到来方向推



定の場合であり，3次元位置推定の場合は検討されていない．よって，反射音に頑健な3次元位置情報の抽出法の開発を行う．

第4章では，残響除去に基づく原音抽出に関する検討を行う．原音をより高詳細に抽出するためには，広帯域なアルゴリズムとして確立する必要があるが，従来法では音声信号を対象とした検討が主であるため広帯域での検討は行われていない．しかし，本研究では，楽器音などのさらに広帯域な信号の原音情報の抽出も必要である．よって，残響除去アルゴリズムの広帯域化の開発を行う．

そして第5章において，本研究の最大の目的である放射指向特性の抽出に関する検討を行う．ここではまず最初に，第4章で抽出した原音情報と観測信号から系のインパルス応答を抽出する．そして，抽出したインパルス応答から放射指向特性の分離方法の開発を行う．

最後に第6章において，本論文の結論を述べる．

## 第2章 包囲型マイクロホンアレイの構築

### 2.1 はじめに

本章では、音空間全体を記録するためのシステムの構築について述べる。1.5 節で述べた通り、記録された音情報から放射指向特性を含む音源の属性を抽出するためには、音源や空間全体を囲む形で音情報を記録する必要がある。そこで、本研究では、室内の天井と壁面 4 面の計 5 面にマイクロホンを格子状に配置して同時収録を行う装置である包囲型マイクロホンアレイを構築する。このシステムは 157 ch と非常に多数のマイクロホンを用いた収録装置であり、157 ch の音信号を同時に記録する技術が求められる。また、第 4 章の室内残響除去に関する検討や第 5 章の放射指向特性の抽出に関する検討では包囲型マイクロホンアレイ 157 ch でのインパルス応答の同時計測が必要となる。

2.2 節において、実験施設として構築した包囲型マイクロホンアレイの仕様について述べる。2.3 節では、構築したシステムの同期確認方法について述べる。2.4 節では、包囲型マイクロホンアレイを用いた室内インパルス応答の測定方法について述べる。2.5 節において本章のまとめを述べる。

### 2.2 包囲型マイクロホンアレイ

音源や空間全体を囲む形で音情報を記録するために、室内の天井と壁面 4 面の計 5 面にマイクロホンを格子状に配置して収録を行う装置である包囲型マイクロホンアレイを構築した。部屋 (防音室) には YAMAHA の AVITECS (特注型) を用いた。マイクロホンには Brüel & Kjær の Type 4951 を 157 本使用した。マイクロホンはアルミパイプを用いて壁面や天井から 30 cm の位置に配置した。マイクロホンの格子間隔は 50 cm とした。構築した包囲型マイクロホンアレイを図 2.1、使用したマイクロホンを図 2.2、室内形状とマイクロホン配置を図 2.3 に示す。このシステムは室内全体を格子状に配置したマイクロホンで囲んでいるため、音源に対して床以外の方向の放射指向特性を含んだ直接音や反射音を記録可能である。

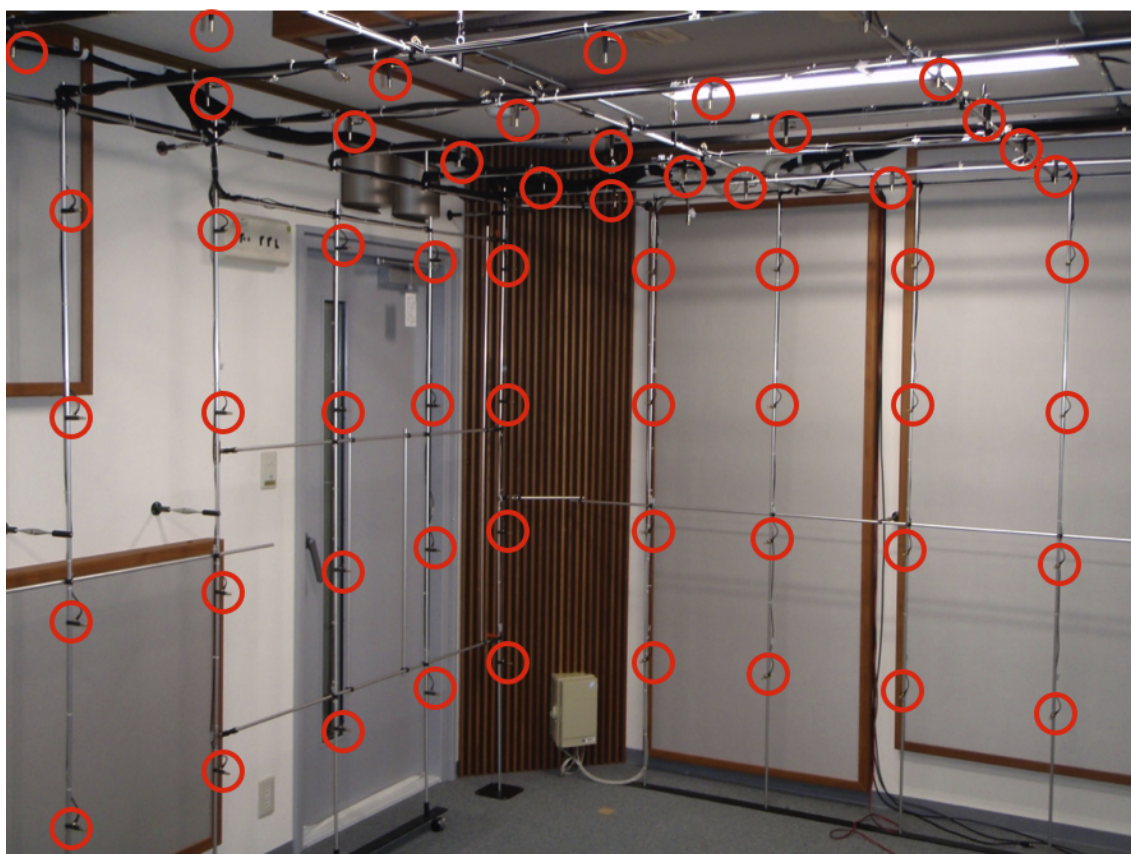


図 2.1: 包囲型マイクロホンアレイの概観      : マイクロホン



図 2.2: Brüel & Kjær Type 4951

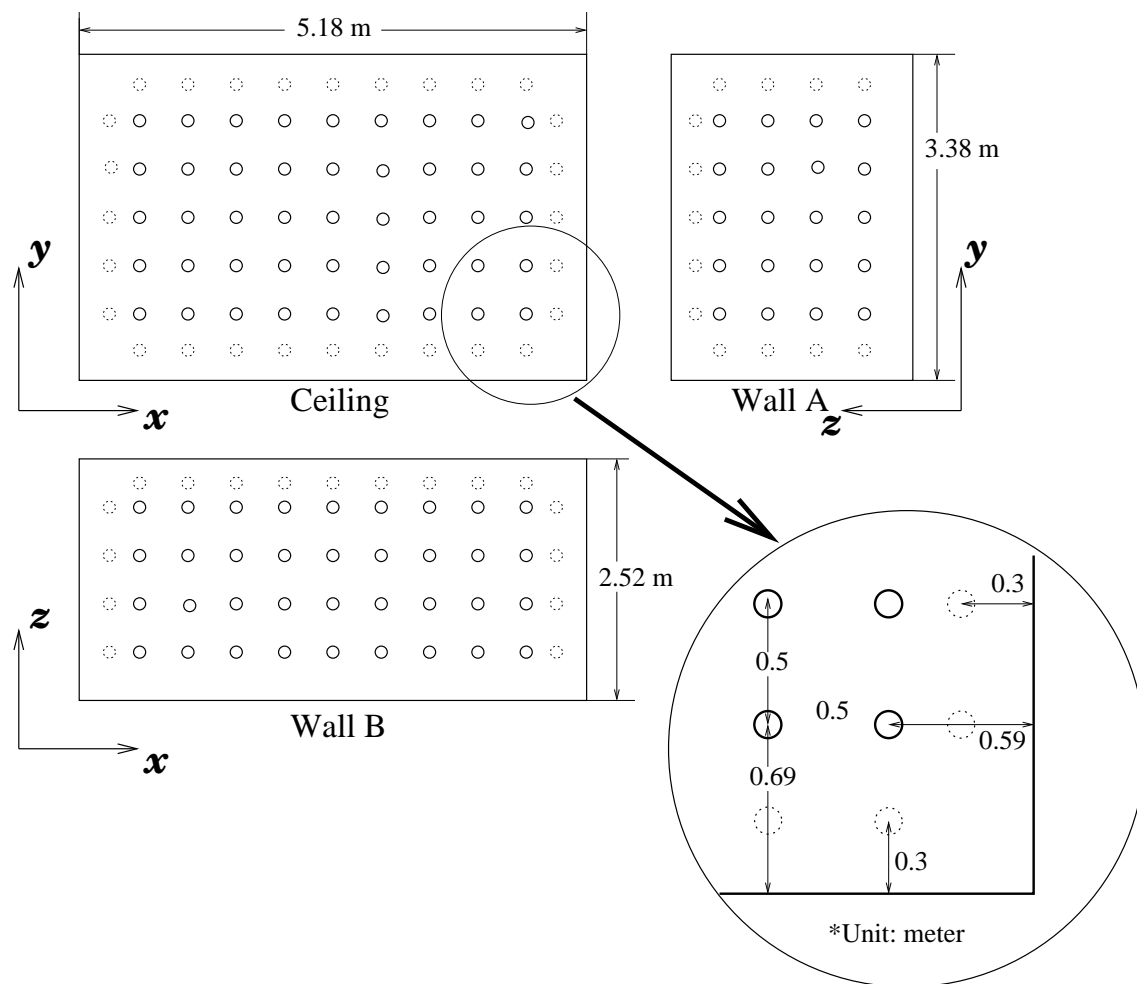


図 2.3: 包囲型マイクロホンアレイのマイクロホン配置

包囲型マイクロホンアレイで収音した音情報を忠実に記録するためには，157 ch の音情報の時間をすべて同期させて記録する必要がある．157 ch という非常に多チャネルの音情報を 1 つの機器を用いて全てを同時に記録することは，計算処理速度やデータの転送，書き込み速度の面から現在の技術では不可能である．そこで本研究では，157 ch の音データを 4 台の PC に分割して記録するという方法を取る．具体的に構築したシステムを図 2.4，使用した機材とそのスペックを表 2.1，制御卓の外観を図 2.5 に示す．

マイクロホンの校正は，Transducer Electric Data Sheet (TEDS) [36, 37] を Brüel & Kjær の DeltaTron 技術によって行う．TEDS とは，センサに内蔵された校正情報を読み取り，自動的に計測設定を行うことのできる IEEE1451.4 [38] に準拠したシステムである．本研究では図 2.4 のように制御 PC とマイクロホンアンプを接続し，制御ソフトである Brüel & Kjær BE 5291 Setup & Control for Type 2694 を用いて 157 ch のマイクロホン情報の読み込みおよび校正の一元管理を行っている．

157 ch をすべて同期させて記録するためには，4 台の PC で使用する A/D・D/A の

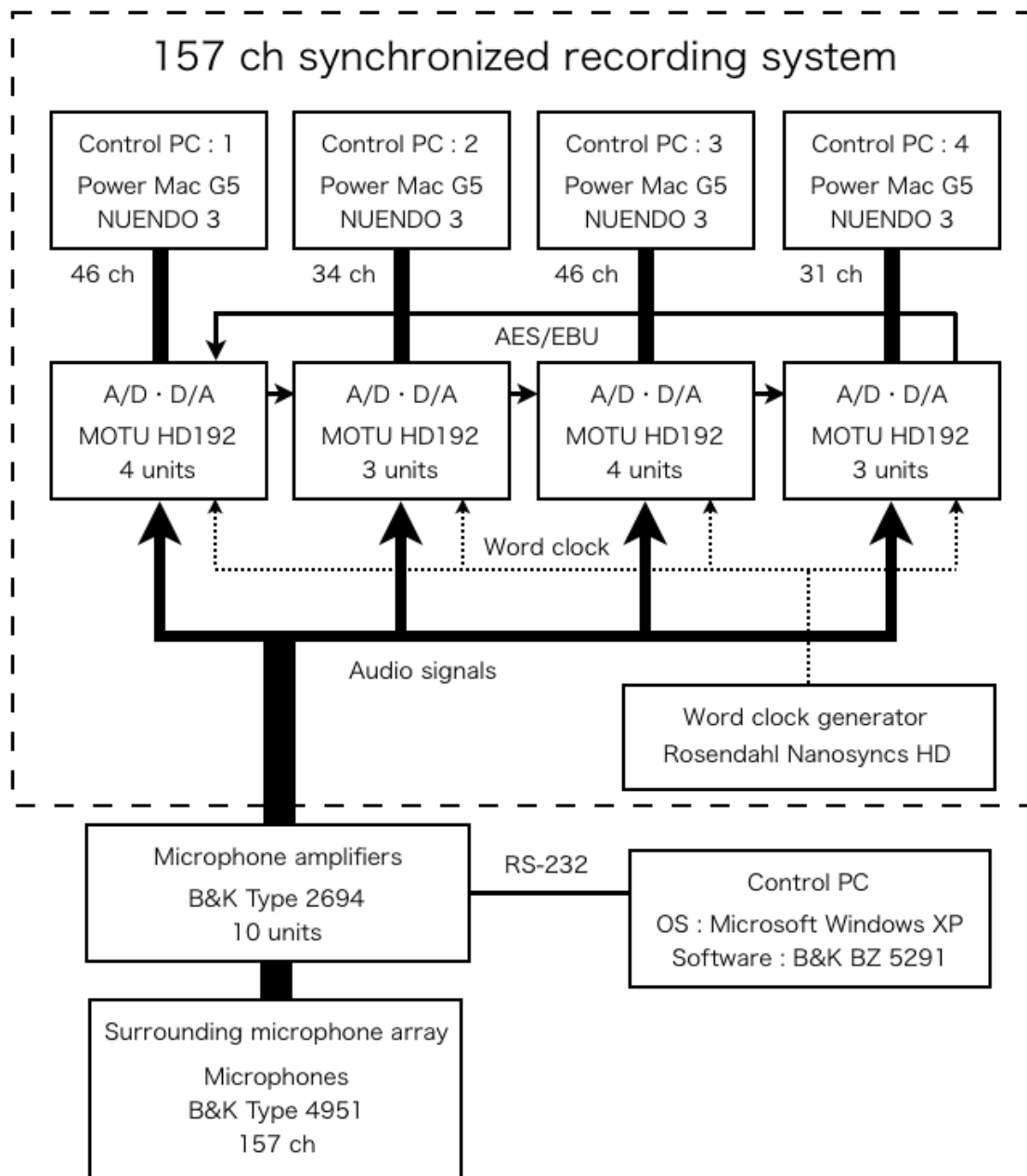


図 2.4: 157 ch 同時収録のシステム構成

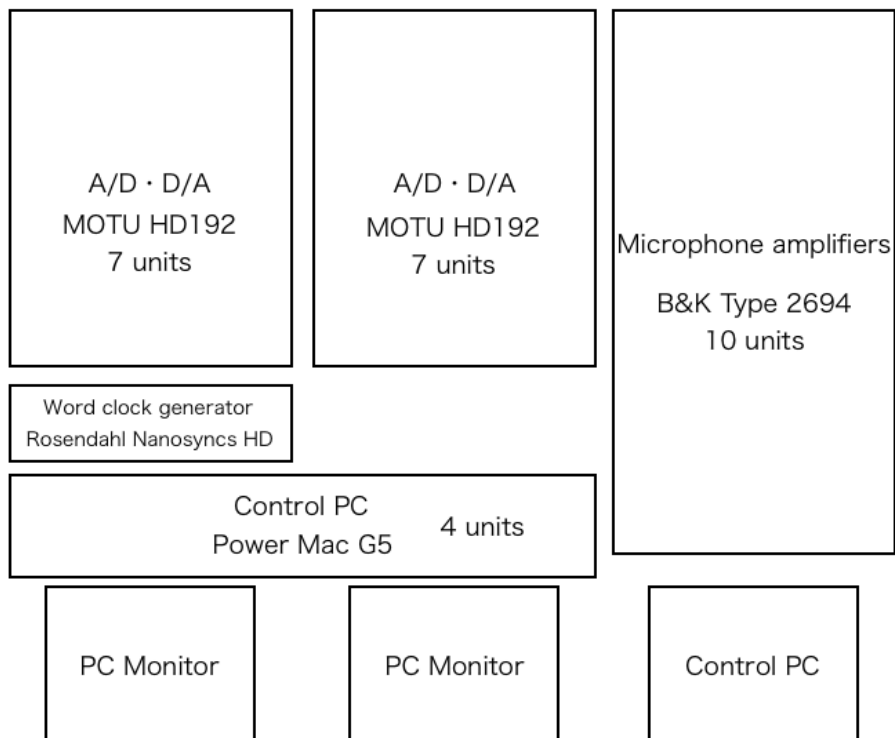


図 2.5: 制御卓の概観

表 2.1: 使用機材とそのスペック

機材	型番	スペック
マイクロホン	Brüel & Kjær Type 4951	感度：6.3 mV/Pa 周波数：10 - 20 kHz ダイナミックレンジ： 30 - 140 dB
マイクロホンアンプ	Brüel & Kjær Type 2694	チャンネル数：16 最大入力電圧：±10 V
A/D & D/A	MOTU HD192	チャンネル数：12 サンプリングレート： 44.1 , 48 , 96 , 192 kHz 量子化ビット：16 , 24 bit
クロックジェネレータ	Rosendahl Nanosyns HD	ワードクロック出力数：8
制御用 PC	Apple Mac Power G5	CPU：PowerPC G5 (3.0) 2.5 GHz Dual メモリ：DDR2 2 GB

クロックおよび収音タイミングを全て同期させる必要がある。クロックの同期には、クロックジェネレータから 48 kHz のワードクロック信号を 4 系統同時に出力し、全ての A/D・D/A がこのクロックを使用することにより実現できる。4 台の PC での収音ソフトウェアには Steinberg の NUENDO 3 を用いる。図 2.4 のように、4 台の PC で 1～46 ch , 47～80 ch , 81～126 ch , 127～157 ch の収音を行う。各マシンでの収音タイミングの同期は Steinberg の VST System Link を用いる。このシステムにより、4 台のマシンを完全に同期させての収音が可能となる。4 台のマシンの通信には Audio Engineering Society/European Broadcasting Union (AES/EBU) [39, 40] を用いる。クロックジェネレータ、A/D・D/A、制御 PC の具体的な接続を図 2.6 に示す。

このシステムにより、サンプリングレート 48 kHz , 16 bit 量子化のリニア PCM の 157 ch 同時収音が可能となる。

## 2.3 157 ch 同時収音の同期確認

2.2 節で構築したシステムを用いて実環境での 157 ch の同時収音を行うためには、157 ch が完全に同期して収音されているかを実験により確認する必要がある。本節では、157 ch 同時収音の同期確認方法と結果について述べる。

同期確認には、Time Stretched Pulse(TSP) 信号 [41–43] を用いた。同期確認のシステム構成を図 2.7 に示す。



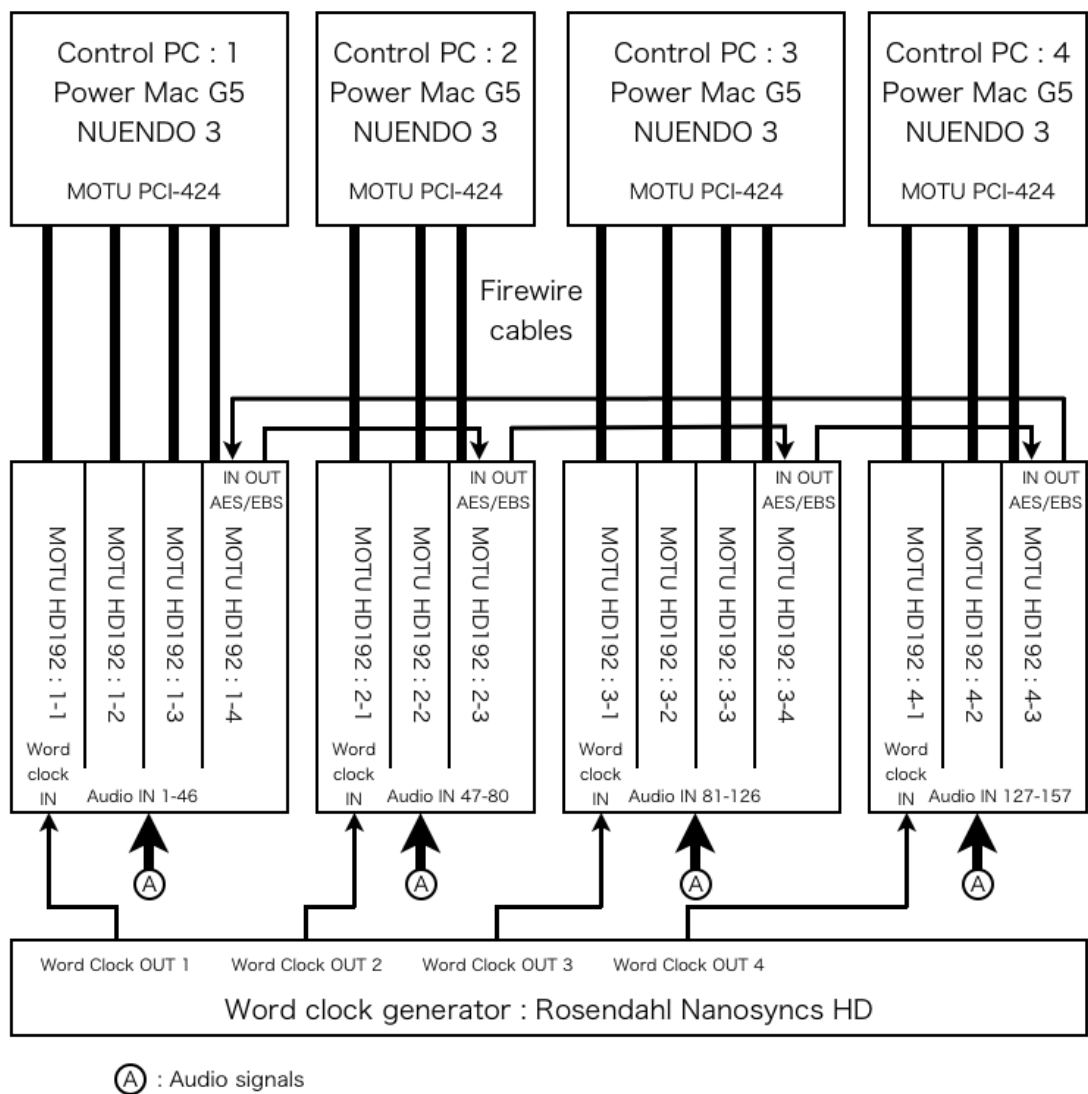


図 2.6: A/D・D/A および各マシンの同期システム



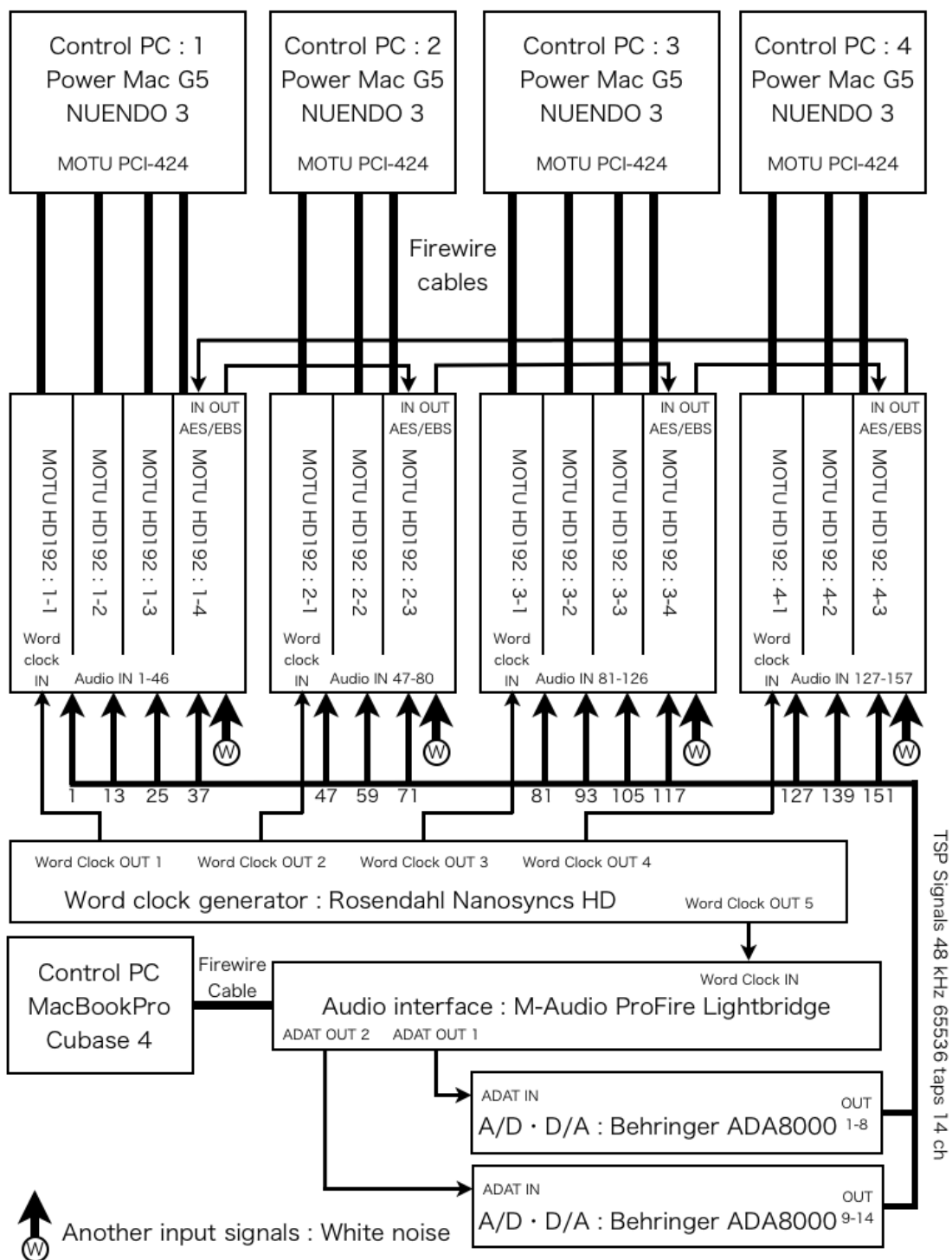


図 2.7: 同期確認作業のシステム構成

オーディオインターフェースと D/A を用いて 65536 ポイント，サンプリングレート 48 kHz の TSP 信号 14 ch を同時に出力し，157 ch 同時収録システムで用いる 14 台の A/D の先頭チャンネルに入力する．この際，入力と出力のクロックも同期させる必要があるため，オーディオインターフェースのクロックも収録システムと同様のクロックジェネレータのワードクロックを用いる．また，この状態では，TSP 信号が入力されているチャンネル以外への入力信号がないため，データ転送や書き込みの条件が 157 ch を同時に収録した場合とは異なる．この状態を防ぐために，その他のチャンネルには包囲型マイクロホンアレイにて出力した白色雑音を入力とする．この作業によって，各 A/D・D/A の先頭チャンネルには TSP 信号，残りのチャンネルには白色雑音が入力され，157 ch 全てに信号が入力される．この作業により収録した 157 ch のデータ (サンプリングレート 48 kHz，16 bit 量子化リニア PCM) から TSP 信号の入力された 14 チャンネルに対して入力に用いた TSP 信号との自己相関を計算する．

以上の手続きを 5 分間連続で収録した結果，全てのチャンネル間でピーク位置が等しく現れることを確認した．これによって，14 台の A/D・D/A はすべて同期して収録されていることが確認できた．また，この作業を各 A/D・D/A ごとに行うことによって，各 A/D・D/A 内での同期も確認できる．

以上により，サンプリングレート 48 kHz，16 bit 量子化のリニア PCM の 157 ch の同時収録の同期を確認した．

## 2.4 インパルス応答測定方法

第 4 章の室内残響除去に関する検討と第 5 章の放射指向特性の抽出に関する検討では，包囲型マイクロホンアレイ 157 ch でのインパルス応答の同時計測が必要となる．インパルス応答の計測には，2.3 節の同期確認でも用いた TSP 信号を用いる．スピーカから TSP 信号を出力し，各観測点で観測した信号に出力した TSP 信号を時間反転させた信号を畳み込むことにより，スピーカと各観測点間のインパルス応答を得ることができる [42]．

TSP 信号を用いたインパルス応答の計測では，スピーカから TSP 信号が出力された時刻と各観測点で収録される信号の時刻の同期が重要となる．本研究では，出力と入力の同期を以下のようなシステムで実現した (図 2.8)．

157 ch の同時収録時には，A/D・D/A のうち 1～157 ch の入力を用いるが，インパルス応答の計測時には，図 2.8 のように 3 ch を入出力に追加して 160 ch を用いての計測を行う．そのうち，158 ch と 160 ch には同じ TSP 信号を出力させる．159 ch では 158 ch の TSP 信号をそのまま受信し，160 ch の TSP 信号をスピーカから出力させる．159 ch で入力される TSP 信号を基準とすることで 1～157 ch の受信信号に対してス

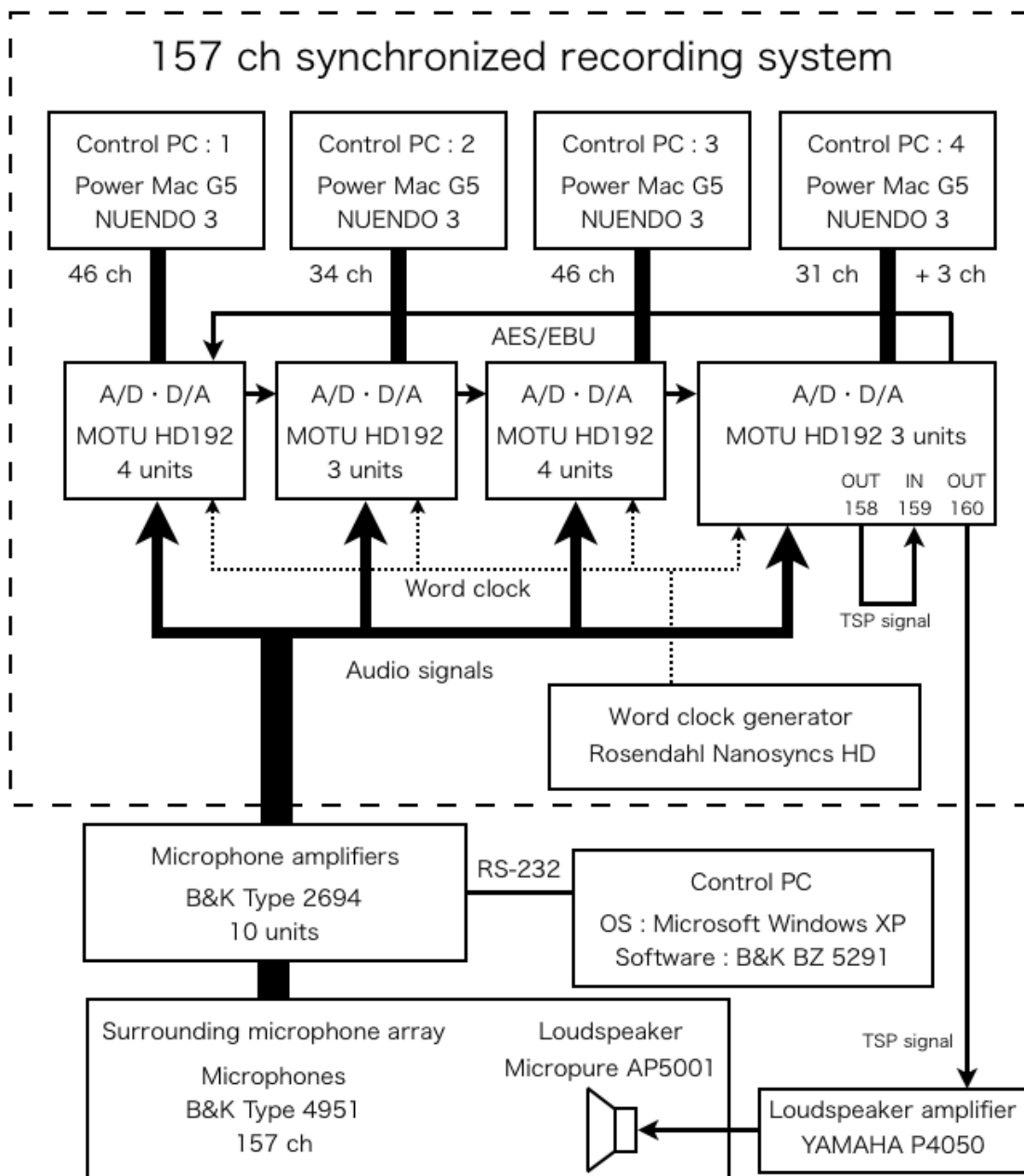


図 2.8: 室内インパルス応答の同時測定

ピーカから出力された TSP 信号の絶対時刻がわかり，時間同期を実現することができる．以上のように収録した 1 ~ 157 ch の信号に時間反転 TSP 信号を畳み込むことによって，157 ch がすべて同期して各観測点でのインパルス応答を得ることが可能となる．

## 2.5 まとめ

本章においては，音空間全体を記録するシステムである包囲型マイクロホンアレイの構築を行った．2.2 節において，実環境に構築した包囲型マイクロホンアレイの仕様について述べた．2.3 節では，構築した収録システムの同期確認の方法について述べた．その結果，サンプリングレート 48 kHz，16 bit 量子化のリニア PCM の 157 ch の同時収録の同期を確認した．2.4 節では，第 4 章および第 5 章の検討で必要となる 157 ch のインパルス応答の計測について述べた．TSP 信号を用いた 157 ch がすべて同期したインパルス応答の計測システムの構築に成功した．

以下，音空間の属性抽出の検討には，本章にて構築を行った包囲型マイクロホンアレイでの観測信号を用いて行うこととする．



## 第3章 反射音に頑健な音源位置推定法の提案

### 3.1 はじめに

本章では，室内における高精度な音源位置推定についての検討を行う．室内における音源位置推定では，直接音と相関の高い反射音が推定精度を低下させるという問題がある．従来法である空間平均化法 [44] では，相関行列をサブアレイを用いて平均化させることにより反射波の影響を低減させることに成功している．しかし，マイクロホンアレイの形状が等間隔直線アレイでない場合や，音波の到来方向推定ではなく音源の位置の推定の場合には，従来手法をそのまま適用することができないという問題がある．そこで本章において，空間平均化法を包囲型マイクロホンアレイに適用する方法について検討する．さらに，入力信号を直接平均化して反射音の影響をさらに低減できるアルゴリズムを提案する．

まず 3.2 節において，マイクロホンアレイを用いた音源位置推定の従来法と，本研究の基礎技術である Multiple Signal Classification (MUSIC) [45] について述べる．3.3 節では，室内における音源位置推定では壁や天井からの反射音によって位置の推定精度が低下するという問題点についてと，空間平均化法を用いて反射音の影響が低減できることについて示す．3.4 節では，音源位置推定かつ不等間隔アレイへの空間平均化法の適用方法について示す．3.5 節では，入力信号を直接平均化して反射音の影響をさらに低減できる Re-Arrangement and Pre-smoothing for MUSIC (RAP-MUSIC) アルゴリズムを提案し，計算シミュレーションと実環境における実験により，従来方法と提案方法の性能評価を行う．3.6 節で本章のまとめを述べる．

### 3.2 マイクロホンアレイを用いた音源位置推定

本節では，マイクロホンアレイを用いた音源位置推定の従来法およびマイクロホン間隔と推定精度について概要を述べる．

### 3.2.1 空間的なフーリエ変換に基づく方法

空間的なフーリエ変換に基づく音波の到来方向推定には，各マイクロホンでの音波の到来時間差に基づく相互相関を用いた方法 [46] や，1.5 節で述べた遅延和などのビームフォーミングの原理に基づく方法 [47–49] などがある．相互相関を用いた方法と遅延和を用いた方法は同義であり，音源到来方向推定の最も基本的な方法である．ここでは，遅延和を用いた方法を例にして説明する．

遅延和アレイは，マイクロホン間の距離に応じた音波の到来時間差に基づき，各マイクロホンの入力信号に遅延を与えて足し合わせることによって，任意の方向に鋭い指向性ビームを向けることができる．アレイの正面方向を  $0 \text{ deg}$  とすると， $\theta_L$  方向に指向性ビームを向けたときのアレイ出力パワー  $p_y(\theta_L)$  は， $\theta_L$  方向から到来する音波の推定値とする．よって，ビーム方向  $\theta_L$  を変化させることにより，方向ごとの到来音パワーが得られる．ビーム方向  $\theta_L$  を変化させた範囲に音源が単一しか存在しない場合には，パワーが最大となった角度  $\theta_{Lmax}$  が音源方向と推定できる．

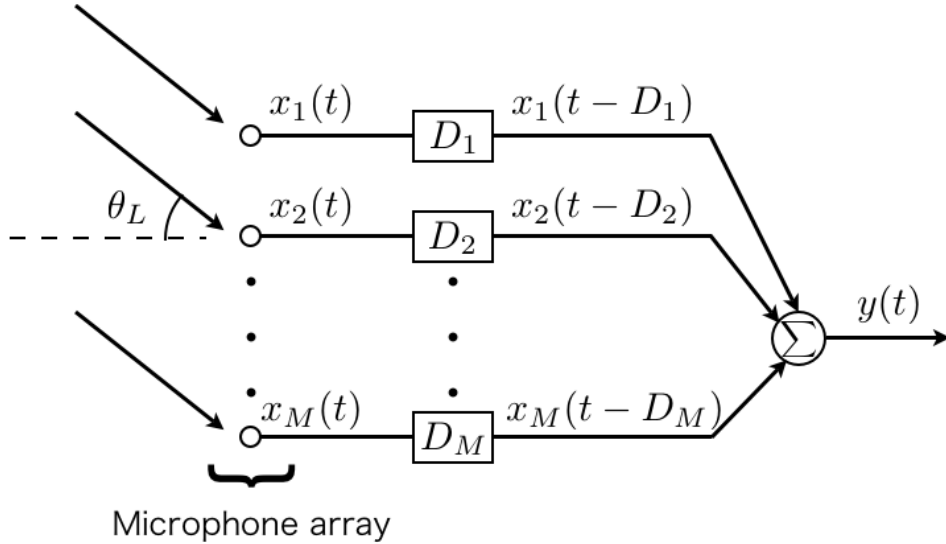


図 3.1: 遅延和アレイ処理のブロック図

$\theta_L$ : 到来方向角,  $x_i(t)$ : 受信信号,  $D_i$ : 遅延器,  $y(t)$ : 出力信号

マイクロホン間隔が  $d$  で，マイクロホン数が  $M$  の直線配列アレイを考える (図 3.1)．角周波数  $\omega$  の複素正弦波が到来しており，各マイクロホンで受音した信号  $x_i(t)$  ( $i = 1, \dots, M$ ) が

$$x_i(t) = X_i e^{j\omega t} \quad (3.1)$$

と表されるものとする． $X_i$  は，各受信信号の複素振幅である．

ここで， $\theta_L$  方向に指向性ビームを向けた遅延和アレイの出力  $y(t)$  は，各受信信号  $x_i(t)$  に各マイクロホン間の時間差  $(i-1)\tau_L$  に基づいて  $e^{j\omega(-D_0(i-1)\tau_L)}$  を乗じて総和をとったものとなる．ここで， $D_i = D_0 - (i-1)\tau$ ， $\tau_L = (d \sin\theta_L)/c$  であり， $D_0$  は遅延量が小さすぎる場合，遅延特性をデジタルフィルタで実現する際の精度が低下することを防ぐために付加する固定遅延量である [50]．よって，アレイ出力パワー  $p_y(\theta_L)$  は，アレイ出力の 2 乗値として次式のように計算される．

$$\begin{aligned} p_y(\theta_L) &= |y(t)|^2 = \left| \sum_{i=1}^M x_i(t) e^{j\omega(-D_0(i-1)\tau_L)} \right|^2 \\ &= \left| \left[ \sum_{i=1}^M X_i e^{j\omega(i-1)\tau_L} \right] e^{j\omega(t-D_0)} \right|^2 \\ &= \left| \sum_{i=1}^M X_i e^{-j(i-1)\Omega_L} \right|^2 \end{aligned} \quad (3.2)$$

$$\Omega_L = -\omega\tau_L = (-\omega d \sin\theta_L)/c \quad (3.3)$$

式 (3.2)，式 (3.3) は，各マイクロホンで観測される複素振幅  $X_1, X_2, \dots, X_M$  を離散時間信号系列と考えれば，DFT(離散フーリエ変換) の式と形式的には同一のものである．つまり，複素振幅系列  $\{X_1, X_2, \dots\} = X_1 \{1, e^{j\Omega_L}, e^{2j\Omega_L}, \dots\}$  は角周波数  $\Omega_L$  の複素正弦波を標本化した信号系列であると考えることができる．このように，遅延和を用いたものは，空間的なフーリエ変換を用いた代表的な音源方向推定方法である．遅延和アレイを利用した方法は，到来する音のパワーの全体的分布を調べるのに適しており，演算量も少ないという特徴を持つ．

しかし，空間的なフーリエ変換を用いた方法にはマイクロホン間隔に依存した空間的折り返し歪みを生じるという欠点もある．ここで，空間のサンプリング周波数  $f_d$  が入力信号の周波数の 2 倍以上でない場合は，離散時間信号の場合と同様，折り返し歪みを生じる．これは，空間的な折り返しひずみと呼ばれ，指向性パターンではサイドローブを形成する．つまり， $0 < \sin\theta_L < 1$  であるため，マイクロホン間隔  $d$  が  $d > c/2f$  ( $f$  は信号の周波数) となる場合は，指向性パターンにサイドローブを生じる．

例えば，本研究で用いる包囲型マイクロホンアレイは  $d = 50$  cm のため， $f > 340$  Hz ではサイドローブが生じる．実際にこの様子を計算機シミュレーションによって確認する．ここでは簡略のため，平面波を用いた音源位置推定によるシミュレーションを行った．図 3.2 のように 8 つのマイクロホンを自由空間に配置した直線アレイ (マイクロホン間隔  $d = 50$  cm) を用いて，2 つの音圧レベルの等しい音源方向 ( $\theta_1 = 0$  deg,  $\theta_1 = 40$  deg) を推定する．図 3.3(a) は音源が共に 300 Hz の場合，図 3.3(b) は音源が共に 800 Hz の場合，図 3.3(c) は音源が共に 3000 Hz の場合である．



マイクロホンを常に狭間隔で配置できる場合は問題ないが，実環境において高い周波数まで保障する間隔でマイクロホンを部屋中に配置することは現実的ではない．

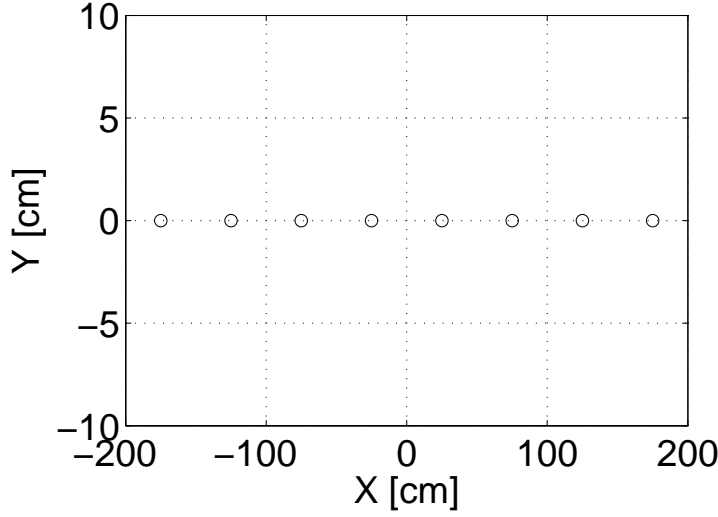


図 3.2: 遅延和と MUSIC の性能評価シミュレーションに用いるマイクロホン配置図

### 3.2.2 MUSIC 法

空間的なフーリエ変換を用いたノンパラメトリックな推定法ではなく，パラメトリックな推定法も提案されている．音の到来方向が (マイクロホン数  $M$ ) $-1$  以下である場合に，音場はこのパラメトリックモデルに適合し，高い分解能の方向推定を行うことができる．このような推定法は高分解能法 (high-resolution method) と呼ばれ，線形予測法 [51]，最小分散法 [51]，MUSIC [45] などがある．その中でも，入力信号の空間相関行列の固有空間構造を用いた MUSIC 法は，パラメトリックな推定法で発生するサイドローブが発生せず，マイクロホン間隔に依存しない高精度な到来方向推定ができる推定法である．以下では，MUSIC の原理について述べる．

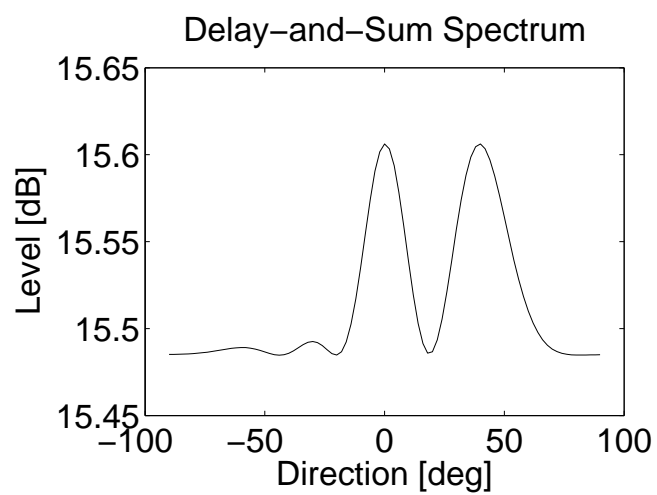
マイクロホンで観測された信号を短区間フーリエ変換したものを以下の入力ベクトル  $\mathbf{x}(\omega, t)$  として定義する．ここで， $\mathbf{A}(\omega)$  は各音源から各マイクロホンへの伝達関数， $s(\omega, t)$  は音源のスペクトル， $\mathbf{n}(\omega, t)$  はマイクロホンで観測される非方向性雑音スペクトルとする．

$$\mathbf{x}(\omega, t) = \mathbf{A}(\omega)s(\omega, t) + \mathbf{n}(\omega, t) \quad (3.4)$$

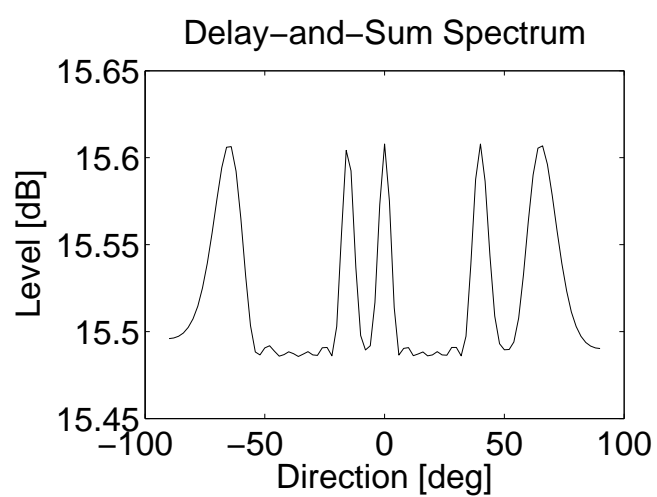
式 (3.4) の第 1 項は方向性音源を，第 2 項は非方向性音源を表す．

空間相関行列  $\mathbf{R}(\omega)$  は  $s(\omega, t)$  と  $\mathbf{n}(\omega, t)$  が無相関であると仮定すると

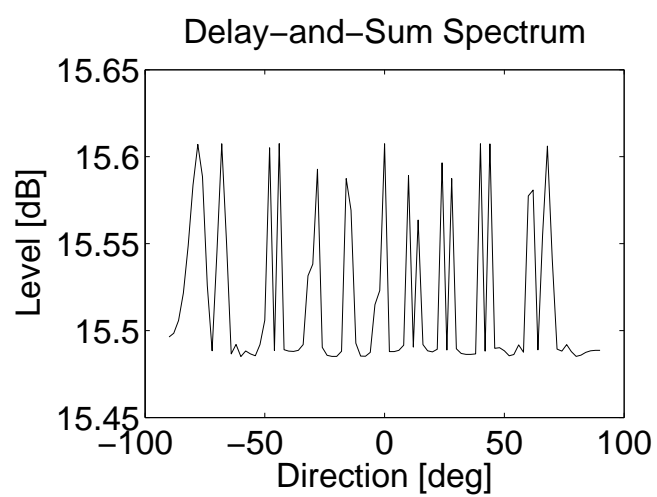
$$\mathbf{R}(\omega) = \text{E}[\mathbf{x}(\omega, t)\mathbf{x}^H(\omega, t)] = \mathbf{A}(\omega)\text{E}[s(\omega, t)s^H(\omega, t)]\mathbf{A}^H(\omega) + \text{E}[\mathbf{n}(\omega, t)\mathbf{n}^H(\omega, t)] \quad (3.5)$$



(a) 300 Hz



(b) 800 Hz



(c) 3000 Hz

図 3.3: マイクロホン間隔 50 cm を用いた場合の遅延和による音源到来方向推定結果

となる．ここで， $E[s(\omega, t)s^H(\omega, t)] = p(\omega)$  は信号の平均パワーであり，また  $n(\omega, t)$  が各マイクロホン間で相関がなく， $E[n(\omega, t)n^H(\omega, t)] = \sigma I$  のモデルで表されるとすると，

$$R(\omega) = A(\omega)p(\omega)A^H(\omega) + \sigma I \quad (3.6)$$

となる．今，音源の数を  $D$  とすると，相関行列  $R(\omega)$  のランクは  $D$  となる．ユニタリ行列  $V$  を用いて  $R(\omega)$  を固有値展開すると，

$$R(\omega) = V \text{diag}(\lambda_1, \dots, \lambda_D, 0, \dots, 0) V^{-1} \quad (3.7)$$

となる．さらに  $VV^H = I$ ， $V^H = V^{-1}$  より， $V$  により  $R(\omega)$  を対角化して，

$$V^H R(\omega) V = [V^H A] p [V^H A]^H = \text{diag}(\lambda_1, \dots, \lambda_D, 0, \dots, 0) + \sigma I \quad (3.8)$$

が得られる．この式から，固有ベクトルを非零の固有値に対応する固有ベクトル ( $V_s = [v_1, \dots, v_D]$  for  $\lambda_1, \dots, \lambda_D$ ) と零の固有値に対応する固有ベクトル ( $V_n = [v_{D+1}, \dots, v_M]$  for  $\lambda_{D+1}, \dots, \lambda_M$ ) に分ける．すると式 (3.8) より，

$$[V_s^H A] p [V_s^H A]^H = \text{diag}(\lambda_1, \dots, \lambda_D) \quad (3.9)$$

$$[V_n^H A] p [V_n^H A]^H = \text{diag}(\sigma, \dots, \sigma) \quad (3.10)$$

が成立する．式 (3.10) から， $V$  と  $A(\omega)$  が直交していることが分かる．ここで， $A(\omega)$  が張る信号部分空間 (signal subspace) に対して， $V_s$  は signal subspace の基底ベクトルとなり， $V_n$  は直交補空間の基底ベクトルとなる．

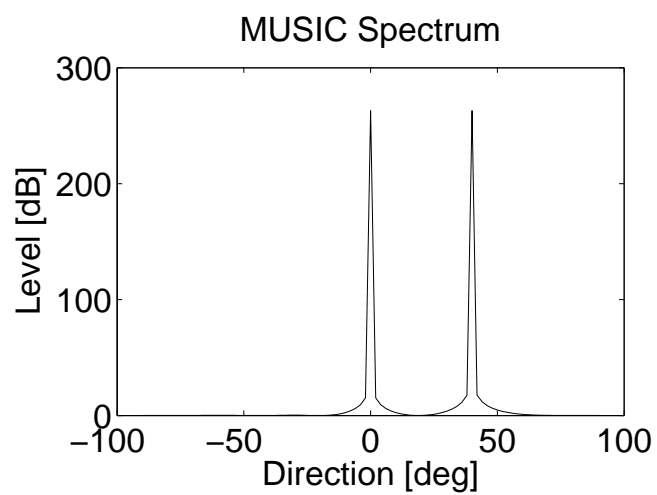
ここで，空間スペクトル  $P$  は，

$$P(\omega, \theta) = \frac{1}{\| V_n^H a(\omega, \theta) \|^2} \quad (3.11)$$

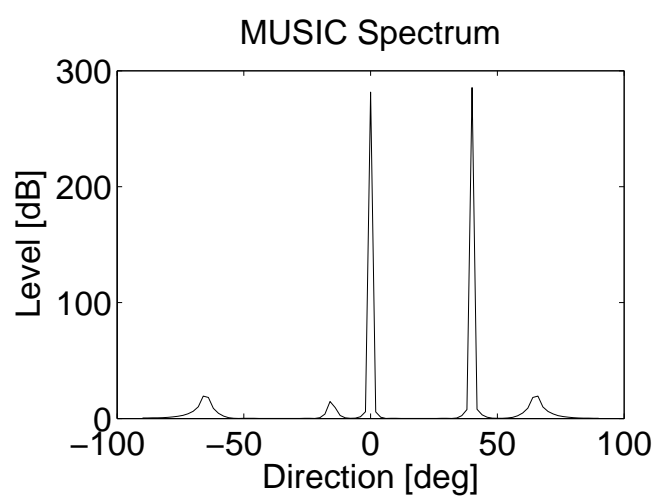
となる．ここで， $a(\omega, \theta)$  は角周波数  $\omega$  における  $\theta$  方向からの伝達関数 (=ステアリングベクトル) であり，音源のありそうな空間全体を走査する．このとき， $a(\omega, \theta)$  が真の音源ベクトルのうちのどれかと一致すれば， $V_n$  とは直交するため，式 (3.11) は大きな値をもつ．

MUSIC 法は，狭帯域信号を対象とした方法であるため，広帯域の信号に用いる場合は，各周波数帯域ごとに MUSIC 法を行い，時間軸上で多数回平均する必要がある．また，到来音のパワーの大きさは推定できないが，到来方向に関しては極めて高い分解能で推定できる [51]．

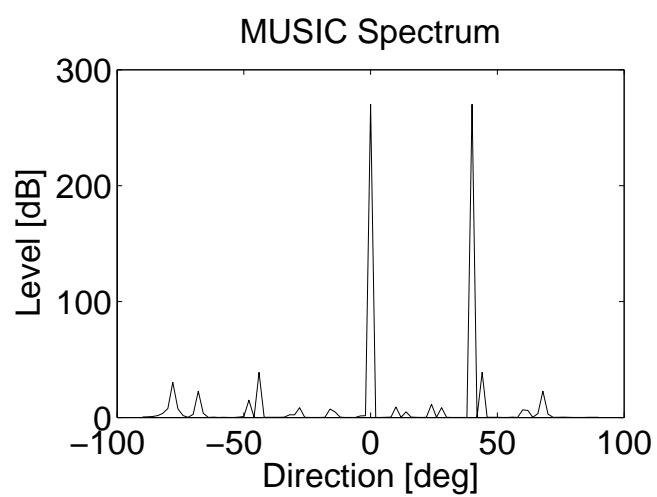
MUSIC アルゴリズムにおける項と同様のシミュレーション結果を図 3.4 に示す．結果より，MUSIC はマイクロホン間隔に依存せず，高周波数まで極めて高い精度で推定できることが確認できる．



(a) 300 Hz



(b) 800 Hz



(c) 3000 Hz

図 3.4: マイクロホン間隔 50 cm を用いた場合の MUSIC による音源到来方向推定結果

### 3.2.3 マイクロホンアレイを用いた音源位置推定

従来のマイクロホンアレイの研究の大半は、到来音波を平面波と見なせる遠距離音場での研究である。理由としては、平面波を仮定した場合、アレイアンテナやデジタルフィルタの設計と共通点が多く、各種の設計方法が確立していることが上げられる。また、使用状況を考えたとき、アレイの大きさに比べて目的音源の位置が遠い場合が多く、アレイ位置ではほぼ平面波として見なすことができる [52–55]。

しかし、包囲型マイクロホンアレイでは、アレイ長が音源距離に対して長いこと、音源の到来方向ではなく位置 (3次元座標) を推定する必要があることの2つの理由により、遠距離音場ではなく近接音場として取り扱う必要がある。

また、アレイ長が音源距離に対して長い場合、マイクロホンと音源との距離は近いものと遠いものでは音源位置によっては  $\sqrt{2}$  倍以上も距離が異なる。マイクロホンアレイと音源の距離が近い近接音場では、到来信号を平面波ではなく球面波として扱わなければならない。本研究で用いる包囲型マイクロホンアレイを考えた場合、マイクロホンアレイが部屋全体を囲っているため、部屋のどこに音源があっても、平面波と見なすことはできない。つまり、本研究においては、到来音波を平面波ではなく球面波として汎用化<sup>はんよう</sup>する必要がある。

到来音波を球面波として扱う場合、時刻  $t$  における音源信号を  $s(t)$  とし、音源から  $i$  番目のマイクロホンまでの距離を  $r_i$ 、音速を  $c$  とすると、距離減衰を考慮することとなるため、 $i$  番目のマイクロホンにおける入力信号  $u_i(t)$  は

$$u_i(t) = \frac{1}{r_i} s(t - r_i/c) \quad (3.12)$$

となる。ここで、音速  $c$  は 340 m/s (気温 15 ) である。

ところで、到来音波を平面波として扱う場合は、音源の到来方向しか探査することができない。しかし、本研究で対象とする音源位置推定は、部屋のどの位置に音源があるかを3次元座標として推定することである。したがって、音源位置を3次元座標として得るためにも、各マイクロホンまでの到来音波を球面波として取り扱う方がよい。到来音波を球面波で取り扱った場合、各マイクロホンでの入力信号は式 (3.12) のようになる。ここで、マイクロホン配置が直線アレイのような1次元配置ではなく、2次元以上の配置であれば、各マイクロホンと探査点との距離の関係は一意に決まるため、音源を3次元的に探査できる。

したがって、平面波仮定におけるステアリングベクトル  $a(\omega, \theta)$  は、式 (3.11) で示されるように角周波数  $\omega$  と入射角  $\theta$  の関数であったが、球面波の場合には、角周波数  $\omega$  と音源位置  $(x, y, z)$  の関数になる。すなわち、点  $(x, y, z)$  と  $i$  番目のマイクロホンにお

けるステアリングベクトル  $\mathbf{a}$  は、3次元探査の場合、

$$a_i(\omega, x, y, z) = \frac{1}{r_i} e^{j\omega \frac{r_i}{c}} \quad (3.13)$$

となる．ここで、 $i$  番目のマイクロホンの座標を  $(x_{mi}, y_{mi}, z_{mi})$  とすると、

$$r_i = \sqrt{(x_{mi} - x)^2 + (y_{mi} - y)^2 + (z_{mi} - z)^2} \quad (3.14)$$

である．また、 $c$  は音速である．

### 3.3 反射音による音源位置推定精度の劣化と空間平均化法

#### 3.3.1 反射音による音源位置推定精度の劣化

ここでは、MUSIC による音源位置推定における反射音の影響について考える．

$L$  本の到来波が全て互いに完全相関 (コヒーレント波) であるとする、第  $l$  到来波の複素振幅は、

$$F_l(t) = \alpha_l F_1(t) (l = 1, 2, \dots, L; \alpha_1 = 1) \quad (3.15)$$

と表される． $\alpha_l$  は、第  $l$  波の第 1 波に対する減衰、位相遅れを表す複素定数である．このときの到来波の複素振幅ベクトル  $\mathbf{F}(t)$  は次式で表される．

$$\begin{aligned} \mathbf{F}(t) &= [F_1(t), F_2(t), \dots, F_L(t)]^T \\ &= [\alpha_1 F_1(t), \alpha_2 F_1(t), \dots, \alpha_L F_1(t)]^T \\ &= F_1(t) [\alpha_1, \alpha_2, \dots, \alpha_L]^T \\ &= F_1(t) \boldsymbol{\alpha} \end{aligned} \quad (3.16)$$

$$\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_L]^T \quad (3.17)$$

マイクロホンアレイが図 3.5 のような  $M$  チャンネルの等間隔直線アレイ (マイクロホン間隔:  $d$ ) であるとする、マイクロホンへの入力ベクトルは内部雑音  $\mathbf{N}(t)$  を含めて、

$$\begin{aligned} \mathbf{X}(t) &= [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \dots, \mathbf{a}(\theta_L)] \begin{bmatrix} F_1(t) \\ F_2(t) \\ \vdots \\ F_L(t) \end{bmatrix} + \mathbf{N}(t) \\ &= F_1(t) \sum_{l=1}^L \alpha_l \mathbf{a}(\theta_l) + \mathbf{N}(t) \\ &= F_1(t) \mathbf{b} + \mathbf{N}(t) \end{aligned} \quad (3.18)$$

と表される．ただし，

$$\mathbf{b} = \sum_{l=1}^L \alpha_l \mathbf{a}(\theta_l) = \mathbf{A} \boldsymbol{\alpha} \quad (3.19)$$

$$\mathbf{A} = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \dots, \mathbf{a}(\theta_L)] \quad (3.20)$$

$$\mathbf{a}(\theta_l) = [1, e^{-j\omega d \sin \theta_l}, \dots, e^{-j\omega(M-1)d \sin \theta_l}]^T \quad (3.21)$$

である．式 (3.18) は，第 1 波のみがあたかも方向ベクトル  $\mathbf{b}$  から到来する表現と等価である．よって，相関行列  $\mathbf{R}_{xx}$  は次式で与えられる．

$$\begin{aligned} \mathbf{R}_{xx} &= \mathbf{E}[\mathbf{X}(t)\mathbf{X}(t)^H] \\ &= \mathbf{A}\mathbf{E}[\mathbf{F}(t)\mathbf{F}(t)^H]\mathbf{A}^H + \sigma^2 \mathbf{I} \\ &= P_1 \mathbf{A} \boldsymbol{\alpha} \boldsymbol{\alpha}^H \mathbf{A}^H + \sigma^2 \mathbf{I} \\ &= P_1 \mathbf{b} \mathbf{b}^H + \sigma^2 \mathbf{I} \end{aligned} \quad (3.22)$$

ただし，

$$\sigma^2 \mathbf{I} = \mathbf{E}[\mathbf{N}(t)\mathbf{N}(t)^H] \quad (3.23)$$

$$P_1 = \mathbf{E}[|\mathbf{F}_1(t)|^2] \quad (3.24)$$

である．

相関行列  $\mathbf{R}_{xx}$  の信号成分を表す行列  $P_1 \mathbf{b} \mathbf{b}^H \equiv \mathbf{R}_{ss}$  の固有値，固有ベクトルを求める．固有値を  $\mu_i$ ，固有ベクトルを  $\mathbf{e}_i$  で表すと，

$$\mathbf{R}_{ss} \mathbf{e}_i = P_1 \mathbf{b} \mathbf{b}^H \mathbf{e}_i = \mu_i \mathbf{e}_i \quad (i = 1, 2, \dots, M) \quad (3.25)$$

と表される．行列  $\mathbf{R}_{ss}$  は非負定値エルミート行列でそのランクは明らかに 1 である．よって，固有ベクトルには

$$\mathbf{e}_i^H \mathbf{e}_k = \delta_{ik} \quad (i, k = 1, 2, \dots, M) \quad (3.26)$$

という関係が成り立ち ( $\delta_{ik}$  はクロネッカーのデルタ)，固有値については

$$\mu_1 > \mu_2 = \dots = \mu_M = 0 \quad (3.27)$$

という関係がある．さらに

$$\mathbf{R}_{ss} \mathbf{b} = (P_1 \mathbf{b} \mathbf{b}^H) \mathbf{b} = P_1 \mathbf{b} (\mathbf{b}^H \mathbf{b}) = P_1 \|\mathbf{b}\|^2 \mathbf{b} \quad (3.28)$$

$$\mathbf{R}_{ss} \frac{\mathbf{b}}{\|\mathbf{b}\|} = P_1 \|\mathbf{b}\|^2 \frac{\mathbf{b}}{\|\mathbf{b}\|} \quad (3.29)$$

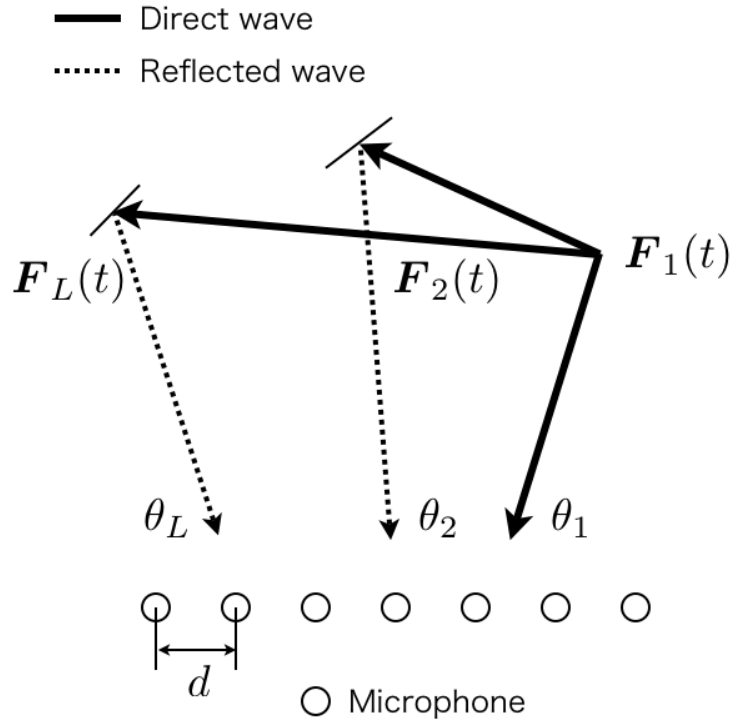


図 3.5: 等間隔直線アレイとコヒーレント波

が成り立つので

$$\mathbf{e}_1 = \frac{\mathbf{b}}{\|\mathbf{b}\|} \quad (3.30)$$

$$\mu_1 = P_1 \|\mathbf{b}\|^2 \quad (3.31)$$

である．

行列  $\mathbf{R}_{ss} \equiv P_1 \mathbf{b} \mathbf{b}^H$  に上記のような関係があるので．相関行列  $\mathbf{R}_{xx} = \mathbf{R}_{ss} + \sigma^2 \mathbf{I}$  の固有値，固有ベクトルには式 (3.25) などにより，次の関係が成り立つ．

$$\begin{aligned} \mathbf{R}_{xx} \mathbf{e}_i &= (\mathbf{R}_{ss} + \sigma^2 \mathbf{I}) \mathbf{e}_i \\ &= \mathbf{R}_{ss} \mathbf{e}_i + \sigma^2 \mathbf{e}_i \\ &= \mu_i \mathbf{e}_i + \sigma^2 \mathbf{e}_i \\ &= (\mu_i + \sigma^2) \mathbf{e}_i \\ &= \lambda_i \mathbf{e}_i \end{aligned} \quad (3.32)$$

よって， $\mathbf{R}_{xx}$  の固有ベクトルは  $\mathbf{R}_{ss}$  のそれと同じとなる，また，固有値  $\lambda_i$  は  $\lambda_i = \mu_i + \sigma^2$  となり，

$$\lambda_1 = \mu_1 + \sigma^2 > \lambda_2 = \dots = \lambda_M = \sigma^2 \quad (3.33)$$



となる．また，式 (3.26) と式 (3.30) より，

$$\mathbf{b}^H \mathbf{e}_i = 0 (i = 2, 3, \dots, M) \quad (3.34)$$

となる．式 (3.33) より，相関行列  $\mathbf{R}_{xx}$  の固有値から到来方向は 1 つと判定されることになる．実際， $L$  本の波が到来しているにも関わらず，それが互いにコヒーレントであるためにそれらの合成波が 1 波とみなされている状態である．したがって，MUSIC 法で到来方向推定を行ったとしても，雑音部分空間に属すると判断された固有ベクトル  $\{\mathbf{e}_2, \mathbf{e}_3, \dots, \mathbf{e}_M\}$  と直交するベクトルは，式 (3.34) よりベクトル  $\mathbf{b}$  ただ一つである． $\mathbf{b}$  は式 (3.19) より， $L$  本の波の方向ベクトルの線形結合である．よって，MUSIC によって真の到来方向を推定することが難しくなる．そこで，この問題を解決する方法として，空間平均化法が提案されている．

### 3.3.2 空間平均化法の原理

アレイアンテナなどでは，コヒーレント波が存在する環境においても所望波の到来方向推定精度を上げるために，所望波と干渉波の相互相関を抑圧する方法が検討されている．相関抑圧法の基本的な考え方は，相関のある波の位相関係（相互相関係数の位相）は受信位置によって異なるので，受信点を平行移動させて相関値を求め，足し合わせた平均を取ると，その平均効果により相互相関値が低下することを利用する．実際には，アレイを動かさずに全体のアレイから同じ配列をする部分（サブアレイ）を複数個取り出し，それらの相関行列を平均する方法を取る．この前処理は空間平均化法（空間平滑化法）[44] と呼ばれ，アレイ構成は同一チャネル数による等間隔直線アレイや長方形平面アレイ，三角形平面アレイなど一定規則で素子配列され，どのサブアレイも平行移動によってのみ重なるアレイに限定される．最も基本的なアレイ構成である等間隔直線アレイを図 3.6 に示す． $M$  チャネル直線アレイから  $K$  チャネルサブアレイを 1 つずつずらしながら合計  $N (= M - K + 1)$  個のサブアレイを取り出す．そして各サブアレイの相関行列（部分相関行列）を適当に重み付けして足し合わせ，所望波と干渉波の相互相関を抑制するのである．アレイに対応する相関行列を図 3.7 に示す．

全アレイの入力ベクトルは，

$$\mathbf{X}(t) = [x_1(t), x_2(t), \dots, x_M(t)]^T \quad (3.35)$$

で示され，第  $n$  サブアレイ ( $K$  チャネル) の入力ベクトル  $\mathbf{X}_n(t)$  は，

$$\mathbf{X}_n(t) = [x_n(t), x_{n+1}(t), \dots, x_{n+K-1}(t)]^T (n = 1, 2, \dots, N) \quad (3.36)$$

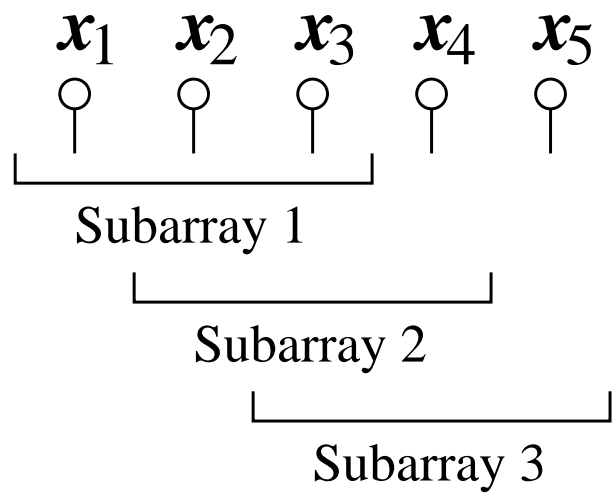


図 3.6: 等間隔直線アレイを用いたサブアレイの構成図

$$\mathbf{R} = \begin{bmatrix}
 \mathbf{R}_{x_1 x_1} & \mathbf{R}_{x_2 x_1} & \mathbf{R}_{x_3 x_1} & \mathbf{R}_{x_4 x_1} & \mathbf{R}_{x_5 x_1} \\
 \mathbf{R}_{x_1 x_2} & \mathbf{R}_{x_2 x_2} & \mathbf{R}_{x_3 x_2} & \mathbf{R}_{x_4 x_2} & \mathbf{R}_{x_5 x_2} \\
 \mathbf{R}_{x_1 x_3} & \mathbf{R}_{x_2 x_3} & \mathbf{R}_{x_3 x_3} & \mathbf{R}_{x_4 x_3} & \mathbf{R}_{x_5 x_3} \\
 \mathbf{R}_{x_1 x_4} & \mathbf{R}_{x_2 x_4} & \mathbf{R}_{x_3 x_4} & \mathbf{R}_{x_4 x_4} & \mathbf{R}_{x_5 x_4} \\
 \mathbf{R}_{x_1 x_5} & \mathbf{R}_{x_2 x_5} & \mathbf{R}_{x_3 x_5} & \mathbf{R}_{x_4 x_5} & \mathbf{R}_{x_5 x_5}
 \end{bmatrix}$$

$\mathbf{R}_1$        $\mathbf{R}_2$        $\mathbf{R}_3$

図 3.7: 3 素子サブアレイの相関行列

で与えられる．したがって，第  $n$  サブアレイの相関行列 (第  $n$  部分相関行列) は次式で表される．

$$\mathbf{R}_{xx} = E[\mathbf{X}_n(t)\mathbf{X}_n(t)^H] (n = 1, 2, \dots, N) \quad (3.37)$$

各部分相関行列に対する重み付けを  $z_n (n = 1, 2, \dots, N)$  とすると  $N$  個の部分相関行列の平均 (空間平均) によって次式の相関行列が得られる．

$$\bar{\mathbf{R}}_{xx} = \sum_{n=1}^N z_n \mathbf{R}_{xx}^n \quad (3.38)$$

ただし，平均の際のパワー保存，および  $\bar{\mathbf{R}}_{xx}$  が正定値エルミート行列であることから，通常  $z_n$  は実数で，次式を満たす．

$$\sum_{n=1}^N z_n = 1 \quad (3.39)$$

MUSIC 法がコヒーレント波の到来方向を正確に推定できないのは，信号相関行列  $\mathbf{S} = E[\mathbf{F}(t)\mathbf{F}(t)^H]$  が，

$$\mathbf{S} = P_1 \boldsymbol{\alpha} \boldsymbol{\alpha}^H \quad (3.40)$$

となってランクが 1 となり，信号部分空間の次元が 1 に縮退してしまうためである．仮に  $L$  本の波が互いに無相関であれば，

$$\mathbf{S} = \begin{bmatrix} P_1 & & \mathbf{O} \\ & \ddots & \\ \mathbf{O} & & P_L \end{bmatrix} \quad (3.41)$$

となって明らかにランクは  $L$  である．したがって，到来方向推定では信号相関行列  $\mathbf{S}$  のランクがフルランク ( $L$ ) に回復することが空間平均導入の直接の目的となる．そのため，部分相関行列の重み付け  $z_n$  を，

$$z_n = \frac{1}{N} (n = 1, 2, \dots, N) \quad (3.42)$$

とし，単純な一様空間平均を用いることが多い．

サブアレイ 1 の入力ベクトルを次のように表す．

$$\mathbf{X}_1(t) = \mathbf{A}\mathbf{F}(t) + \mathbf{N}_1(t) \quad (3.43)$$

$$\mathbf{A} = [\mathbf{a}(\theta_1), \mathbf{a}(\theta_2), \dots, \mathbf{a}(\theta_L)] \quad (3.44)$$

$$\mathbf{a}(\theta_l) = [1, e^{-j\omega d \sin \theta_l}, \dots, e^{-j\omega (M-1)d \sin \theta_l}] (l = 1, 2, \dots, L) \quad (3.45)$$

ただし， $\mathbf{N}_1(t)$  はサブアレイ 1 の内部雑音ベクトルである，このとき全てのサブアレイの入力ベクトルは次式のように表される．

$$\mathbf{X}_n(t) = \mathbf{A}\mathbf{B}^{n-1}\mathbf{F}(t) + \mathbf{N}_n(t) (n = 1, 2, \dots, N) \quad (3.46)$$

ここで，行列  $B$  は次式で定義される  $L$  次の対角行列である．

$$\begin{aligned} B &= \text{diag}(\nu_1, \nu_2, \dots, \nu_L) \\ &= \begin{bmatrix} \nu_1 & 0 & \cdots & 0 \\ 0 & \nu_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \nu_L \end{bmatrix} \end{aligned} \quad (3.47)$$

$$\nu_l = e^{-j\omega d \sin \theta_l} \quad (l = 1, 2, \dots, L) \quad (3.48)$$

また， $N_n(t)$  は第  $n$  サブアレイの内部雑音ベクトルである．よって，第  $n$  サブアレイの相関行列  $R_{xx}$  は次式で表される．

$$\begin{aligned} R_{xx}^n &= E[X_n(t)X_n(t)^H] \\ &= AB^{n-1}E[F(t)F(t)^H](B^{n-1})^H A^H + E[N_n(t)N_n(t)^H] \\ &= AB^{n-1}S(B^{n-1})^H A^H + \sigma^2 I \end{aligned} \quad (3.49)$$

空間平均として一様平均空間法の場合，得られる相関行列は次式となる．

$$\begin{aligned} \bar{R}_{xx} &= \frac{1}{N} \sum_{n=1}^N R_{xx}^n \\ &= A \left[ \frac{1}{N} \sum_{n=1}^N B^{n-1} S (B^{n-1})^H \right] A^H + \sigma^2 I \\ &= A \bar{S} A^H + \sigma^2 I \end{aligned} \quad (3.50)$$

ただし，

$$\bar{S} = \frac{1}{N} \sum_{n=1}^N B^{n-1} S (B^{n-1})^H \quad (3.51)$$

$K$  チャネルのサブアレイ 1 だけの場合は，信号相関行列  $S$  は式 (3.40) で示され， $\text{rank}(S) = 1$  である．それに対し，空間平均によって得られる相関行列の信号相関行列  $\bar{S}$  は式 (3.51) に式 (3.40) を代入することにより次式のように表される．

$$\begin{aligned} \bar{S} &= \frac{1}{N} \sum_{n=1}^N B^{n-1} (P_1 \alpha \alpha^H) (B^{n-1})^H \\ &= \frac{P_1}{N} \sum_{n=1}^N (B^{n-1} \alpha) (B^{n-1} \alpha)^H \\ &= \frac{P_1}{N} [\alpha, B\alpha, B^2\alpha, \dots, B^{N-1}\alpha] \begin{bmatrix} \alpha^H \\ (B\alpha)^H \\ (B^2\alpha)^H \\ \vdots \\ (B^{N-1}\alpha)^H \end{bmatrix} \\ &= \frac{P_1}{N} \mathbf{C} \mathbf{C}^H \end{aligned} \quad (3.52)$$

ここで ,

$$\begin{aligned}
\mathbf{C} &= [\alpha, B\alpha, B^2\alpha, \dots, B^{N-1}\alpha] \\
&= \begin{bmatrix} \alpha_1 & \nu_1\alpha_1 & \nu_1^2\alpha_1 & \cdots & \nu_1^{N-1}\alpha_1 \\ \alpha_2 & \nu_2\alpha_2 & \nu_2^2\alpha_2 & \cdots & \nu_2^{N-1}\alpha_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha_L & \nu_L\alpha_L & \nu_L^2\alpha_L & \cdots & \nu_L^{N-1}\alpha_L \end{bmatrix} \\
&= \mathbf{D}\mathbf{V}
\end{aligned} \tag{3.53}$$

ただし ,

$$\mathbf{D} = \begin{bmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \alpha_L \end{bmatrix} \tag{3.54}$$

$$\mathbf{V} = \begin{bmatrix} 1 & \nu_1 & \nu_1^2 & \cdots & \nu_1^{N-1} \\ 1 & \nu_2 & \nu_2^2 & \cdots & \nu_2^{N-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \nu_L & \nu_L^2 & \cdots & \nu_L^{N-1} \end{bmatrix} \tag{3.55}$$

式 (3.52) より , 行列  $\bar{\mathbf{S}}$  のランクは  $L \times N$  の行列  $\mathbf{C}$  のランクに等しい . すなわち ,  $\text{rank}(\bar{\mathbf{S}}) = \text{rank}(\mathbf{C})$  である . また , 式 (3.53) から ,  $\mathbf{C} = \mathbf{D}\mathbf{V}$  で , 行列  $\mathbf{D}$  は  $L \times L$  の正則行列のため ,  $\text{rank}(\mathbf{C}) = \text{rank}(\mathbf{V})$  である . さらに , 行列  $\mathbf{V}$  は  $L \times N$  の vandermonde 行列のため ,  $N \geq L$  であるならば ,  $\text{rank}(\mathbf{V}) = L$  となる [61] . したがって ,

$$\text{rank}(\bar{\mathbf{S}}) = \text{rank}(\mathbf{C}) = \text{rank}(\mathbf{V}) = L \tag{3.56}$$

が得られる . このように , 空間平均後の信号相関行列  $\bar{\mathbf{S}}$  のランクは  $L$  となり , フルランクまで回復する . この状態で  $R_{xx}$  の固有値や固有ベクトルを求めれば , 到来波が無相関のときと同様に , MUSIC 法においても到来方向を推定することができる .

なお , 正確な推定を行うために必要な条件 (ランク回復条件) としては

$$N = M - K + 1 \geq L \tag{3.57}$$

であり , さらに MUSIC 法の適応条件として

$$K \geq L + 1 \tag{3.58}$$

である . よって , 式 (3.57) , 式 (3.58) より

$$M = N + K - 1 \geq L + K - 1 \geq 2L \tag{3.59}$$

となり , 到来音源数の 2 倍以上のチャネル数が必要となる [56] .

文献 [57–60] などでは空間平均化法を用いて従来法よりも精度のよい到来方向推定を行っている .

### 3.3.3 不均一アレイへの応用

上記の方法では，アレイが等間隔直線配列でこれを構成する観測点配置のパターンが同一であることが前提となっている．したがって，円型アレイのように配列が等間隔でないアレイには空間平均化法をそのまま適用することはできない．

そこで不規則アレイに対しては，受信信号を補間処理により等間隔直線アレイ配列の仮想アレイの受信信号に変換した後に，空間平均化法を用いる方法が提案されている [62–64]．不規則アレイから仮想直線アレイへの補間処理としては，到来波の入射が予想される角度範囲で両アレイのステアリングベクトルを比較し，最小 2 乗法で補間行列を決定する方法が示されている．この方法では，不規則アレイの配列が仮想の等間隔直線アレイのそれと大きく異なる場合や，観測点配置パターンのばらつきが大きい場合，及び測角範囲が広い場合には，補間処理の誤差が増大し，測角精度が大きく劣化する問題がある．

入射波の到来が予想される角度範囲中の互いに異なった  $P$  個の角度ポイント  $\Theta_{\text{all}} \equiv \{\theta_1, \theta_2, \dots, \theta_P\}$  において，不規則アレイのステアリングベクトル  $\mathbf{a}$  で構成される行列を  $\mathbf{A}(\Theta_{\text{all}})$ ，仮想等間隔直線アレイのステアリングベクトル  $\bar{\mathbf{a}}$  で構成される行列を  $\bar{\mathbf{A}}(\Theta_{\text{all}})$  とする． $\mathbf{A}(\Theta_{\text{all}})$ ，および  $\bar{\mathbf{A}}(\Theta_{\text{all}})$  は，それぞれ次式で定義される  $M \times P$  行列である．

$$\mathbf{A}(\Theta_{\text{all}}) = [\mathbf{a}(\theta_1) \cdots \mathbf{a}(\theta_P)] \quad (3.60)$$

$$\bar{\mathbf{A}}(\Theta_{\text{all}}) = [\bar{\mathbf{a}}(\theta_1) \cdots \bar{\mathbf{a}}(\theta_P)] \quad (3.61)$$

ここで， $\bar{\mathbf{a}}(\theta)$  は次式のようになる．

$$\bar{\mathbf{a}}(\theta) = [1, \eta(\theta), \eta(\theta)^2, \dots, \eta(\theta)^{M-1}]^T \quad (3.62)$$

$\eta(\theta)$  は仮想直線アレイの配列で決まる関数であり， $\bar{d}$  を仮想直線アレイの素子間隔とすれば， $\eta(\theta) = e^{-j\omega\bar{d}\sin\theta}$  で与えられる．

補間行列は次式を満足する行列である．

$$\bar{\mathbf{A}}(\Theta_{\text{all}}) = \mathbf{B}\mathbf{A}(\Theta_{\text{all}}) \quad (3.63)$$

一般には  $M < P$  となるため，式 (3.63) は不定であり直接解くことはできない．そこで，式 (3.63) を最小 2 乗問題ととらえ，次式の評価関数  $\Psi_{\text{all}}$  を最小にする  $\mathbf{B}$  を補間行列  $\mathbf{B}'$  として用いる．

$$\Psi_{\text{all}}(\mathbf{B}) = \|\mathbf{B}\mathbf{A}(\Theta_{\text{all}}) - \bar{\mathbf{A}}(\Theta_{\text{all}})\|_F^2 \quad (3.64)$$

ここで， $\|\mathbf{Q}\|_F^2 = \text{trace}(\mathbf{Q}^H \mathbf{Q})$  と定義される．

式 (3.64) で求めた不規則アレイから仮想等間隔直線アレイへの補間行列  $B'$  を用いて、受信信号の共分散行列  $R$  を次式のように変換する．

$$\bar{R}' = B' R B'^H = \bar{A}' R_S \bar{A}'^H + \sigma^2 B' B'^H \quad (3.65)$$

$$\bar{A}' = B' A \quad (3.66)$$

$R_S$  は入力信号  $s(t)$  の共分散行列 ( $R_S = E[s(t)s(t)^H]$ ) である．そして、 $\bar{R}'$  を次式により空間平均化し、

$$\hat{\bar{R}}' = \frac{1}{N} \sum_{n=1}^N \bar{R}'_n \quad (3.67)$$

$\hat{\bar{R}}'$  を用いて MUSIC 法を行う．ここで、 $\bar{R}'_n$  は第  $n$  サブアレイの相関行列で、 $N$  は空間平均回数 (=サブアレイの数) である．

式 (3.64) の補間誤差  $\Psi_{\text{all}}$  が十分小さくなるような  $B'$  が求められれば、 $\bar{A}' \cong \bar{A}$  となり、 $\bar{R}'$  は等間隔直線アレイにおける受信信号相関行列と見なせる．このような場合、第  $i$  サブアレイのステアリングベクトルで構成される行列  $\bar{A}_i$  が

$$\bar{A}_i = \bar{A}_j \mathcal{D}_{i-j} \quad (3.68)$$

$$\bar{\mathcal{D}} = \text{diag}(\eta, (\theta_1), \eta(\theta_2), \dots, \eta(\theta_k)) \quad (3.69)$$

の関係を満足するため、コヒーレントな信号が入射する場合であっても、 $R_s$  が空間平均化されてフルランクになる．

しかし、この方法では、不規則アレイの配列や素子パターンが等間隔直線アレイのそれらと大きく異なる場合、補間誤差  $\Psi_{\text{all}}$  が一般に大きくなる．その結果、式 (3.68) が成立しなくなるため、到来方向推定精度が劣化する．

この問題に対して、文献 [62] では補間処理の誤差を抑圧し測角度精度を確保する方法として、 $\Theta_{\text{all}}$  を単純に縮小する方法が提示されている．しかし、この方法では、測角精度が劣化する恐れがある． $\Theta_{\text{all}}$  を縮小し過ぎて一部の入射波の到来角が  $\Theta_{\text{all}}$  の外になると、式 (3.68) が成立しないため、精度が劣化するためである．これらの問題を解決する方法として、相関行列の変換を段階的に行う段階的補間 MUSIC (Stepped Interpolation SSP MUSIC) 法 [66] や、その前段階の大まかな測角に AVW 法 (Adding Virtual Waves 法) [66] を用いた方法 [67] が提案されている．

### 3.4 空間平均化法の音源位置推定への適用

アレイアンテナでは、3.3 節で述べたように空間平均化法を用いることにより、コヒーレント波が入射してくる状況においても高精度に MUSIC 法を行うことができる．

したがって本研究においても，空間平均化法を行うことによって，コヒーレント波(＝反射音)のある環境においても高精度な音源位置推定が期待できる．

3.3 節で述べた空間平均化法は，到来波の方向推定であって音源位置を推定しているわけではない．一方，本研究においては包囲型マイクロホンアレイによる 3 次元音源位置推定が必要であるため，空間平均化法をそのまま適用することはできない．したがって，本節では，3 次元位置推定に空間平均化法を適用させるための方法について検討する．

### 3.4.1 音源位置推定への適用の問題点

空間平均化法は，平均化に用いる相関行列の空間的なマイクロホン配置が全て等しく，かつ等間隔配置における方向推定のために，多重化した後の各サブアレイのステアリングベクトルも全て等しいことを利用している方法である．これは式 (3.63) の変換行列  $B$  を用いて変換された不均一アレイの仮想等間隔直線アレイでも同様である．

本研究で用いる包囲型マイクロホンアレイは，どの壁面のアレイも格子状を成している．したがって，空間平均化法において空間的に等しいマイクロホン配置を重ね合わせるができるという点については問題ない．

ここで，このアレイを用いて音源の到来方向だけではなく，到来位置を推定する場合を考える．この場合，探査位置から各マイクロホンまでの距離は不均一になるため，ステアリングベクトルも式 (3.62) のように表すことができない．よって，アレイをサブアレイに分けて多重化しても，多重化された各チャンネルのステアリングベクトルは不均一のため，空間平均化法をそのまま適用することができない．

しかし，マイクロホンの情報を多重化して所望波とコヒーレント波の相関を低減する考え方自体は，本研究においても効果を発揮することが期待できる．そこで，この考え方を 3 次元位置推定においても使用できるように拡張することを考える．

### 3.4.2 3 次元位置推定における空間平均化法の原理

ある 1 点の式 (3.11) における  $P$  を計算する場合の平均化は，文献 [62–64] のように各マイクロホンの入力信号を空間平均化できるように仮想チャンネルの信号に変換し，仮想チャンネルにおいて空間平均化法を行えばよい．

球面波の仮定の下で，実チャンネルから仮想チャンネルへの変換は，振幅補正と位相変化によって行う．例えば，図 3.8 の実チャンネルの信号を仮想チャンネルの信号と見なすためには，走査位置から実チャンネルまでの距離を  $r_i$ ，走査位置から仮想チャンネルまでの距離を  $r'_i$ ，時刻  $t$  における実チャンネルの信号を  $x(t)$  とすると，時刻  $t$  における仮想チャ



ネルの信号  $x'(t)$  は，球面波の仮定により，

$$x'(t) = \frac{r}{r'} x(t) e^{-j\omega(\frac{r-r'}{c})} \quad (3.70)$$

の変換によって生成される．この変換によって，仮想チャネルと探査位置間のステアリングベクトルを MUSIC 法において使用することができる．

仮想チャネルの具体的な作成方法を図 3.8 に示す．図 3.8 の上側 4 つのマイクロホンから構成されるサブアレイを実チャネルセット 1，右側 4 つのマイクロホンから構成されるサブアレイを実チャネルセット 2 とする．その上で，それぞれのセットのマイクロホン信号を仮想チャネルでの信号に変換する際，探査点と仮想チャネル間の距離がセット間で等しくなるような位置にそれぞれの仮想チャネルを作成する．これによって仮想チャネルセット 1 と仮想チャネルセット 2 のステアリングベクトルは全て同じになるため，それぞれの仮想チャネルセットで求めた相関行列を用いて空間平均化法を適用することが可能となる．

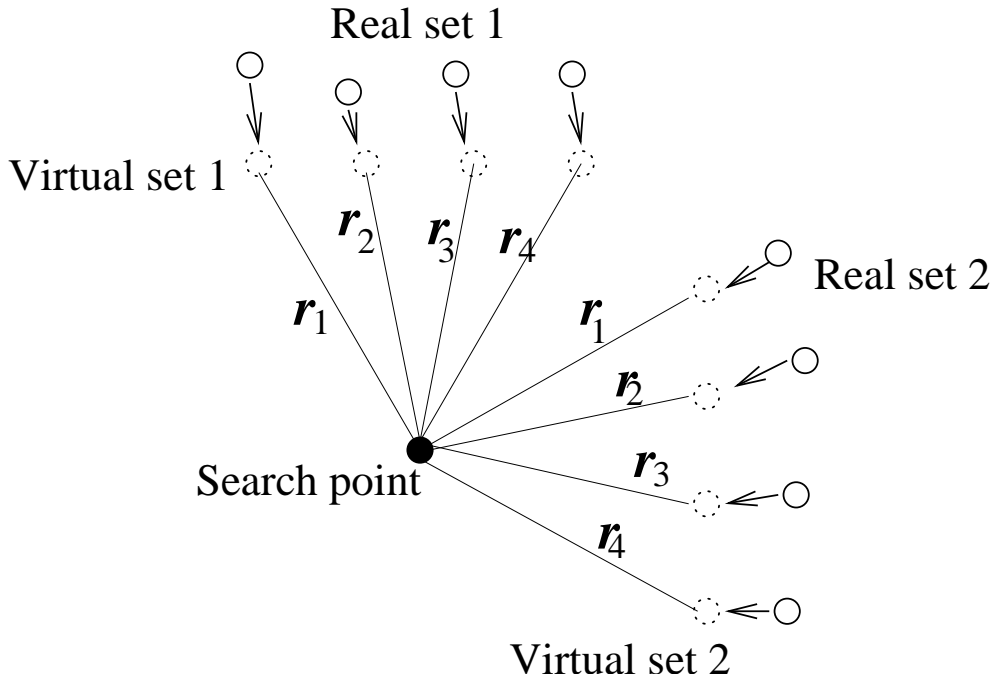


図 3.8: 仮想チャネルセット 1 と仮想チャネルセット 2 のステアリングベクトルが等しくなるための仮想チャネル配置図

従来の空間平均化法の考え方は，MUSIC 法の際に，各サブアレイと探査方向の同じステアリングベクトルが全て同じであることを利用している．方向推定の場合は，どの探査方向においてもアレイ配置が等間隔ならば，ステアリングベクトルは探査方向の角度  $\theta$  の関数になるため，どの方向においてもそのままの相関行列を使用することができる．しかし，3 次元位置探査においては，探査点を移動させると仮想チャネルと

探査位置の距離は場所に応じて変化するため、空間平均化するために各仮想サブアレイを重ね合わせることはできない。移動させた探査点においても空間平均化するためには、また新たに平均化できるように仮想チャネルを作り直さなければならない。よって、3次元位置推定においては、仮想チャネルを探査点ごとに式(3.70)によって生成し、その信号を用いて探査点ごとに相関行列を計算しなければならない。

したがって、3次元位置推定において空間平均化法を行うためには、探査点と各仮想チャネルセットの距離が探査点ごとに同じになるように(＝空間平均できるように)仮想チャネルを生成し直し、各探査点ごとの相関行列を用いて MUSIC 法を行えばよい。

しかし、これは文献[65]でいう探査範囲をただ1点に減らしたものになるため、相関行列のランクが回復せず、音源位置を正しく推定できない危険性がある。そこで、3次元位置推定における空間平均化法を、ランク回復以外の観点で考える。

本研究では、反射音が存在する環境においても、精度よく直接音の位置を推定することを目標としている。空間平均化法においてランクが回復するということは、相関行列の直接音が形成する固有値に反射音の要素が入り込まず、反射音はその固有値を形成することを意味する。しかし、本研究において想定する環境下では、反射音は無数に存在するため、反射音の位置推定は難しい。また、必ずしも反射音の位置推定は必要ではない。よって、本研究においては、相関行列をフルランクに回復することよりも、直接音の形成する固有値に反射音の要素が入り込まないことが重要である。

3次元位置推定において仮想チャネルセットの相関行列を式(3.38)によって空間平均化する場合、多重化する場合に対応する各チャネルの信号は探査位置を基準として全て同相化されている。探査点に音源  $S(t)$  があり、別の位置に干渉波の音源  $C(t)$  がある場合、マイクロホン  $i$  での信号は、その他の雑音を無視し、マイクロホン  $i$  と音源および干渉波の音源との距離をそれぞれ  $r_{Si}$ ,  $r_{Ci}$ , 音速を  $c$  とすると、到来時間はそれぞれ  $\tau_{Si} = r_{Si}/c$ ,  $\tau_{Ci} = r_{Ci}/c$  となるため、

$$x_i(t) = \frac{1}{r_{Si}} S(t) e^{-j\omega\tau_{Si}} + \frac{1}{r_{Ci}} C(t) e^{-j\omega\tau_{Ci}} \quad (3.71)$$

となる。よって、仮想チャネルにおいては、音源と仮想チャネルとの距離を  $r'_S$ , 到来時間を  $\tau'_{Si} = \frac{r'_{Si}}{c}$  として式(3.70)で変換し、

$$x'(t) = \frac{r_{Si}}{r'_{Si}} x_i(t) e^{-j\omega(\tau'_{Si} - \tau_{Si})} = \frac{1}{r'_{Si}} S(t) e^{-j\omega\tau'_{Si}} + \frac{r_{Si}}{r'_{Si} r_{Ci}} C(t) e^{-j\omega(\tau_{Ci} - \tau_{Si} + \tau'_{Si})} \quad (3.72)$$

となる。

ここで、空間平均化した場合としない場合とでは、MUSIC 法においてどの程度違いが出るかを相関行列  $R$  を見ることによって検証する。空間平均化しない場合、サブア

レイ  $n$  ( $n = 1, 2, \dots, N$ ) の相関行列の  $(p, q)$  成分  $R(p, q)$  は式 (3.71) より ,

$$\begin{aligned}
R(p, q) &= E[x_p(t)x_q(t)^*] \\
&= \frac{1}{r_{Sp}r_{Sq}} P_S e^{-j\omega(\tau_{Sp}-\tau_{Sq})} \\
&\quad + \frac{1}{r_{Sp}r_{Cq}} E[S(t)C(t)^*] e^{-j\omega(\tau_{Sp}-\tau_{Cq})} \\
&\quad + \frac{1}{r_{Cp}r_{Sq}} E[C(t)S(t)^*] e^{-j\omega(\tau_{Cp}-\tau_{Sq})} \\
&\quad + \frac{1}{r_{Cp}r_{Cq}} P_C e^{-j\omega(\tau_{Cp}-\tau_{Cq})}
\end{aligned} \tag{3.73}$$

となる . ここで ,  $P_S = E[S(t)S(t)^*]$  ,  $P_C = E[C(t)C(t)^*]$  である .

空間平均する場合 , サブラレイでの  $R$  は式 (3.72) より ,

$$\begin{aligned}
R'_n(p, q) &= E[x'_{p+n-1}(t)x'_{q+n-1}(t)^*] \\
&= \frac{1}{r'_{Sp+n-1}r'_{Sq+n-1}} P_S e^{-j\omega(\tau'_{Sp+n-1}-\tau'_{Sq+n-1})} \\
&\quad + \frac{r_{Sq+n-1}}{r'_{Sp+n-1}r'_{Sq+n-1}r_{Cq+n-1}} E[S(t)C(t)^*] \\
&\quad \cdot e^{-j\omega(\tau'_{Sp+n-1}-\tau_{Cq+n-1}+\tau_{Sq+n-1}-\tau'_{Sq+n-1})} \\
&\quad + \frac{r_{Sp+n-1}}{r'_{Sp+n-1}r'_{Sq+n-1}r_{Cq+n-1}} E[C(t)S(t)^*] \\
&\quad \cdot e^{-j\omega(\tau_{Cp+n-1}-\tau_{Sp+n-1}+\tau'_{Sp+n-1}-\tau'_{Sq+n-1})} \\
&\quad + \frac{r_{Sp+n-1}r_{Sq+n-1}}{r'_{Sp+n-1}r_{Cp+n-1}r'_{Sq+n-1}r_{Cq+n-1}} P_C \\
&\quad \cdot e^{-j\omega(\tau_{Cp+n-1}-\tau_{Sp+n-1}+\tau'_{Sp+n-1}-\tau_{Cq+n-1}+\tau_{Sq+n-1}-\tau'_{Sq+n-1})}
\end{aligned} \tag{3.74}$$

となる . ここで ,  $E_q(n)$  と  $E_p(n)$  を ,

$$E_q(n) = \frac{r_{Sq+n-1}}{r_{Cq+n-1}} e^{j\omega(\tau_{Sq+n-1}-\tau_{Cq+n-1})} \tag{3.75}$$

$$E_p(n) = \frac{r_{Sp+n-1}}{r_{Cp+n-1}} e^{j\omega(\tau_{Cp+n-1}-\tau_{Sp+n-1})} \tag{3.76}$$

とすると , 式 (3.74) は ,

$$\begin{aligned}
R'_n(p, q) &= \frac{1}{r'_{Sp+n-1}r'_{Sq+n-1}} \left\{ P_S e^{-j\omega(\tau'_{Sp+n-1}-\tau'_{Sq+n-1})} \right. \\
&\quad + E[S(t)C(t)^*] e^{-j\omega(\tau'_{Sp+n-1}-\tau'_{Sq+n-1})} E_q(n) \\
&\quad + E[C(t)S(t)^*] e^{-j\omega(\tau'_{Sp+n-1}-\tau'_{Sq+n-1})} E_p(n) \\
&\quad \left. + P_C e^{-j\omega(\tau'_{Sp+n-1}-\tau'_{Sq+n-1})} E_q(n)E_p(n) \right\}
\end{aligned} \tag{3.77}$$

となる．式 (3.77) を式 (3.38) に代入し，空間平均化法を適用する．この際，探査点と各仮想チャネルとの距離は  $r_{Sp} = r_{Sp+1} = \dots = r_{Sp+N-1}$ ， $r_{Sq} = r_{Sq+1} = \dots = r_{Sq+N-1}$  と等しくなっている．よって，空間平均化適用後の空間相関行列を  $\bar{R}'(p, q)$  とすると，

$$\begin{aligned}
\bar{R}'(p, q) &= \sum_{n=1}^N \frac{1}{N} R'_n(p, q) \\
&= \frac{1}{r'_{Sp} r'_{Sq}} P_S e^{-j\omega(\tau'_{Sp} - \tau'_{Sq})} \\
&\quad + \frac{1}{r'_{Sp} r'_{Sq}} E[S(t)C(t)^*] e^{-j\omega(\tau'_{Sp} - \tau'_{Sq})} \cdot \frac{1}{N} \sum_{n=1}^N E_q(n) \\
&\quad + \frac{1}{r'_{Sp} r'_{Sq}} E[C(t)S(t)^*] e^{-j\omega(\tau'_{Sp} - \tau'_{Sq})} \cdot \frac{1}{N} \sum_{n=1}^N E_p(n) \\
&\quad + \frac{1}{r'_{Sp} r'_{Sq}} P_C e^{-j\omega(\tau'_{Sp} - \tau'_{Sq})} \cdot \frac{1}{N} \sum_{n=1}^N E_q(n) E_p(n) \tag{3.78}
\end{aligned}$$

となる．平均化適用前の式 (3.73) と適用後の式 (3.78) を比較すると，空間平均化した場合の方が第 2 項から第 4 項の効果によって干渉波の影響を低減できていることが分かる．

この方法では，探査位置が音源位置に一致した時には式 (3.78) の効果が現れるため，反射音の影響を低減し，MUSIC 法における式 (3.11) の  $P$  は大きな値を持つことが考えられる．しかし，探査位置に音源が存在しない場合は，式 (3.78) が効果を発揮することがないため， $P$  は大きな値にはならない．つまり，この方法は各点において別々の相関行列を用いているため，空間平均だけによって相関行列のランクを上げるのではなく，直接音の同相化の効果と相関行列の平均化の両方によって干渉波の影響を低減して直接音の位置推定を高精度化する方法であると言える．よって，各点において別々の相関行列を用いて探査しても問題ないことが分かる．

## 3.5 RAP-MUSIC

### 3.5.1 仮想信号の平均化

前述で導入した信号処理法では，同相化と相関行列の多重化によって干渉波の影響を低減している．しかし，各仮想チャネルの信号が同相化しているならば，相関行列を平均化するのではなく，仮想チャネルの信号そのものを始めから平均化し，平均化した信号を用いて相関行列を求める方法も考えられる．これは，遅延和の原理に相当する．つまり，MUSIC 法を行う前に，仮想チャネルのサブアレイで遅延和を行い，その信号を用いて MUSIC 法を行う．この場合， $K$  チャネルで構成されるサブアレイで

同相化された信号の和  $\bar{x}'(t)$  は式 (3.72) より ,

$$\begin{aligned}
\bar{x}'(t) &= \frac{1}{K} \sum_{n=1}^N x'(t) \\
&= \frac{1}{r'_{Si}} S(t) e^{-j\omega\tau'_{Si}} \\
&\quad + \sum_{n=1}^N \frac{1}{K} \frac{r_{Si+n-1}}{r'_{Si} r_{Ci+n-1}} C(t) e^{-j\omega(\tau_{Ci+n-1} - \tau_{Si+n-1} + \tau'_{Si})}
\end{aligned} \tag{3.79}$$

となる . よって , 同相化された信号  $\bar{x}'(t)$  から形成される空間相関行列は

$$\begin{aligned}
\bar{R}'(p, q) &= E[x'_p(t)x'_q(t)^*] \\
&= \frac{1}{r'_{Sp}r'_{Sq}} P_S e^{-j\omega(\tau'_{Sp} - \tau'_{Sq})} \\
&\quad + \frac{1}{r'_{Sp}r'_{Sq}} E[S(t)C(t)^*] e^{-j\omega(\tau'_{Sp} - \tau'_{Sq})} \cdot \frac{1}{N} \sum_{n=1}^N E_q(n) \\
&\quad + \frac{1}{r'_{Sp}r'_{Sq}} E[C(t)S(t)^*] e^{-j\omega(\tau'_{Sp} - \tau'_{Sq})} \cdot \frac{1}{N} \sum_{n=1}^N E_p(n) \\
&\quad + \frac{1}{r'_{Sp}r'_{Sq}} P_C e^{-j\omega(\tau'_{Sp} - \tau'_{Sq})} \cdot \frac{1}{KN} \sum_{n=1}^N E_q(n) \sum_{n=1}^N E_p(n)
\end{aligned} \tag{3.80}$$

となる . 空間平均化法を適用した式 (3.79) と比較すると , 式 (3.80) の第 4 項が  $\frac{1}{KN}$  となっていることから , この推定法の方がより反射音の影響を低減できる .

### 3.5.2 RAP-MUSIC の提案

3.5.1 で示した遅延和处理と MUSIC 法を組み合わせた信号処理法を具体的に実装する方法について示す .

仮想チャネルは , 多重化する際に探査点との距離が対応する仮想チャネルにおいて全て等しいという条件を満たせばよい . 仮想チャネルの位置や組み合わせは無数に存在する . ここで , できるだけ離れたマイクロホン同士を多重化した方が反射音の抑圧には有効であることが考えられる . しかし , 音源が点音源でない場合 , 音源の放射特性を崩してしまう . つまり , 直接音と回折音を区別なく平均してしまうことになるため , 音源の放射特性をできるだけ保つよう , 近接のマイクロホン同士での同相化を行うことを検討した . そこで , 本研究においては , 以下のようにして仮想チャネルを生成する .

まず , 図 3.9 のように 4 つの正方形のマイクロホンを 1 つのセットと見なし , その重心を仮想チャネルとする . そして , この 4 つのマイクロホンの入力信号を式 (3.70) で仮想チャネルにおける信号に変換する . すると , この 4 つの信号は探査点からの信号

が仮想チャンネルにおいて同相化される．よって，これらの組み合わせを図 3.5 のように全ての格子状の重心について行うことにより，仮想チャンネルを生成できる．その上で，それぞれの仮想チャンネルを平均化することによって，仮想チャンネルにおいて遅延処理が適用されたことになる．最後に，各仮想チャンネルで平均化した信号を用いて MUSIC 法を行い，これを各探索点について行う．本提案法を RAP-MUSIC (Re-arregement And Pre-smoothingfor MUSIC) と呼ぶ．

RAP-MUSIC で用いている仮想チャンネルの原理は文献 [68] でも用いられているが，音源の到来方向推定にはビームフォーミングを用いており，高周波数においてはサイドローブの発生が見られる．一方，RAP-MUSIC では，最初に遅延和アルゴリズムを用いて反射音成分を低減させ，反射音成分が低減された信号を用いて MUSIC を行う．つまり，音源位置推定のアルゴリズムはあくまでも空間相関行列の固有空間構造を利用している．したがって，最初に遅延和を用いているが，音源位置推定自体には空間的なフーリエ変換を用いていないため，サイドローブが発生することはない．つまりこの提案法は遅延和と MUSIC の長所のみを使用した方法であると言える．

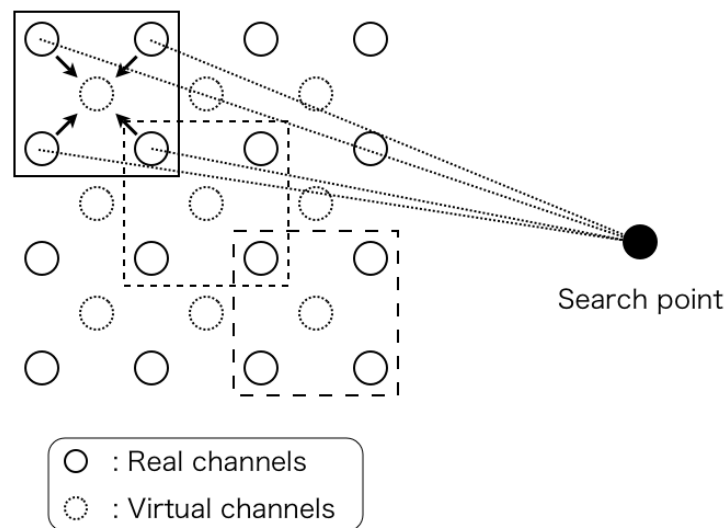


図 3.9: 4 チャンネルの重心位置に仮想チャンネルを作成する概念図

### 3.5.3 実験による提案法の性能評価

3.4 節で考案した音源位置推定への空間平均化法の適用，さらに 3.5.2 項で考案した RAP-MUSIC が実際に反射音の存在する環境においても高精度に音源位置を推定できるかについて計算機シミュレーションおよび実環境における実験により評価を行う．

### 3.5.4 実環境における実験条件

実環境で収録した信号を用いた性能評価を行う．部屋の形状，大きさ，音源位置を図 3.10 に示す．使用するマイクロホンは図 3.11 のように壁 3 面の合計 45 個を用いる．「・」が実チャンネル (45 チャンネル) を表し，「□」が仮想チャンネル (30 チャンネル) を表す．音源のスピーカには図 3.12 に示す直径 7 cm の正 32 面体点音源スピーカ [69] を用いる．このスピーカは，スピーカから 25 cm 以上離れた地点では点音源と見なすことができることが測定により確認されている [69] ．

MUSIC，空間平均化法，RAP-MUSIC の 3 方法についての性能評価を行う．使用する音源はホワイトノイズ，男性音声，女性音声，ギター音を用いた．音声については NTT アドバンステクノロジ株式会社の親密度別単語理解度試験用音声データベースを使用し，ギター音は EBU (European Broadcasting Union) の試験用音源である SQAM (Sound Quality Assessment Material) の一節 (2 s) を用いた．実験条件を表 3.1，3.2 に示す．3.2 に示す通り，パターン 1 から 3 は 1 音源における音源位置推定であり，パターン 4 と 5 については 2 音源における音源位置推定である．

表 3.1: 実験条件

サンプリング周波数	44100 Hz
信号長	1 s
解析フレーム長	512 ポイント
窓関数	ハミング窓
解析周波数帯域	500 ~ 8000 Hz

表 3.2: 音源の種類と提示レベル

セッション名	音源 1 (dB)	音源 2 (dB)
セッション 1	ホワイトノイズ 64.1	
セッション 2	男性音声 63.5	
セッション 3	ギター音 63.0	
セッション 4	ホワイトノイズ 64.1	女性音声 57.9
セッション 5	男性音声 63.5	女性音声 50.4

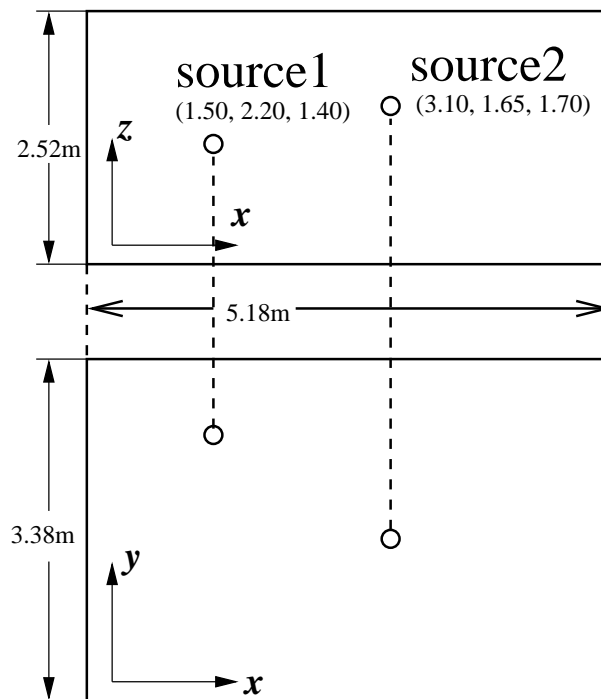


図 3.10: 音源位置推定実験における音源の位置関係

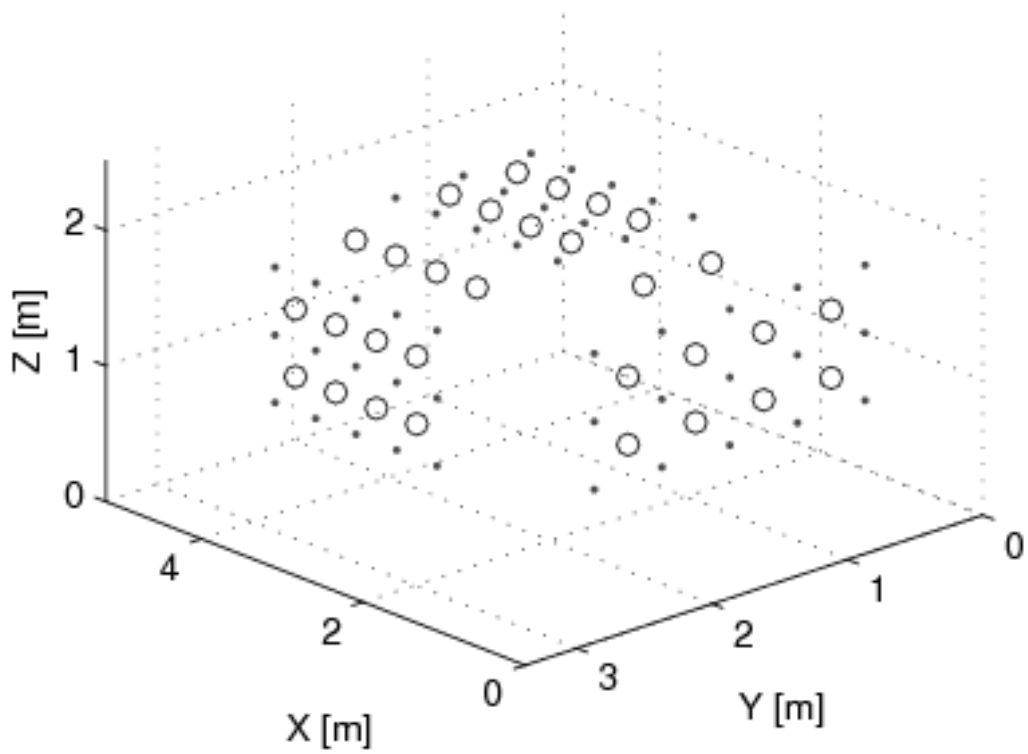


図 3.11: 実験に使用するマイクロホン (3 面) 「・」: 実チャネル, 「○」: 仮想チャネル



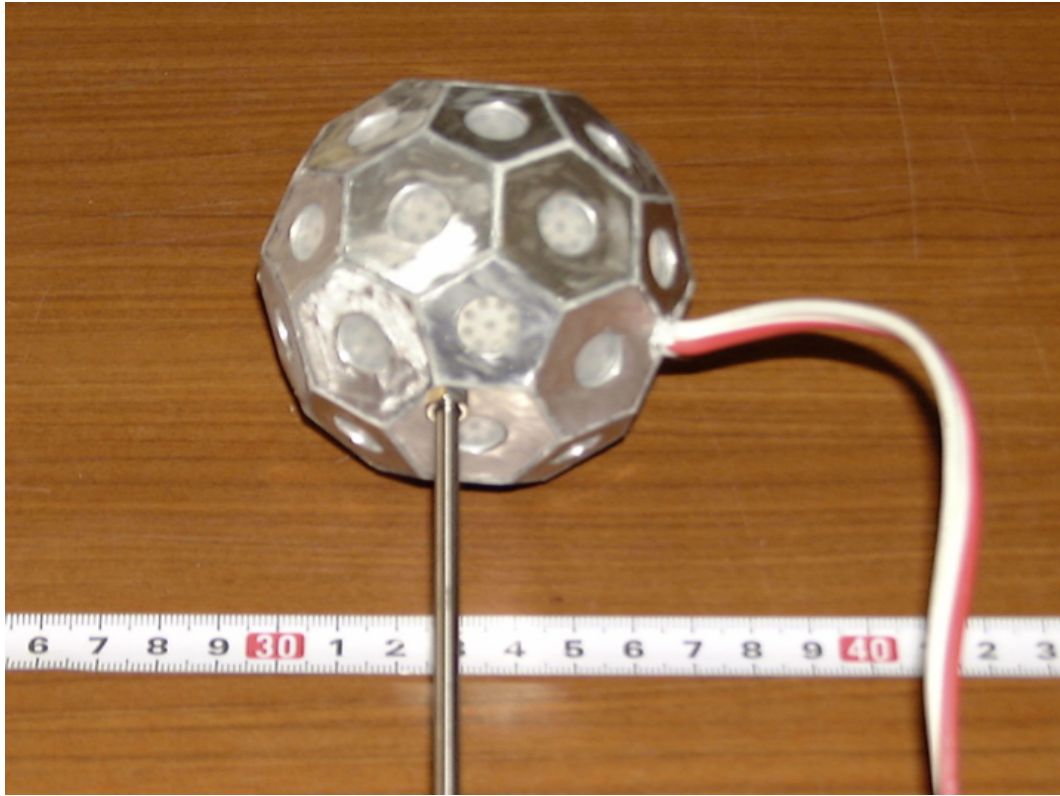


図 3.12: 32 面体点音源スピーカ

### 3.5.5 実環境における実験結果

MUSIC 法は，3.2.2 項で述べた通り，式 (3.11) の  $P$  がピークを持つ位置に音源があると推定する方法である．また， $P$  の値は到来音波のパワーには比例しない．よって，推定した音源位置での  $P$  の値が，いかに他の位置に比べて大きな値であるかを見るために，次式で定義される  $k$  という値を導入する．

$$k = \frac{\text{各探査位置における } P}{\text{設定した音源位置における } P} \quad (3.81)$$

つまり，音源位置では  $k = 1$  となり，他の探査位置では  $k$  が 0 に近い値であるほど音源位置において  $P$  が鋭いピークを形成していることを意味する．表 3.3～表 3.7 に，各セッションにおける各パターンの推定結果 (上位 5 つずつ) を示す．

表 3.3: セッション 1 (ホワイトノイズ) の推定結果

(a) MUSIC

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
1.50	2.20	1.40	10.60	1.0
1.50	2.20	1.50	5.62	0.530
1.50	2.20	1.30	5.60	0.528
1.60	2.20	1.40	5.30	0.500
1.60	2.10	1.30	5.14	0.485

(b) 空間平均化法

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
1.50	2.20	1.40	9.52	1.0
1.50	2.30	1.50	5.29	0.556
1.50	2.30	1.40	5.26	0.553
1.60	2.20	1.40	5.22	0.548
1.60	2.10	1.30	5.14	0.540

(c) RAP-MUSIC

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
1.50	2.20	1.40	16.60	1.0
1.50	2.20	1.50	5.33	0.321
1.50	2.20	1.30	5.05	0.304
1.60	2.20	1.40	4.21	0.254
1.50	2.10	1.40	3.97	0.239

表 3.4: セッション 2 (男性音声) の推定結果

(a) MUSIC

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
1.50	2.20	1.40	7.29	1.0
1.50	2.20	1.50	3.88	0.533
1.60	2.10	1.30	3.70	0.508
1.50	2.20	1.30	3.65	0.501
1.60	2.20	1.40	3.50	0.480

(b) 空間平均化法

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
1.50	2.20	1.40	6.63	1.0
1.60	2.10	1.30	4.16	0.628
1.50	2.30	1.50	3.69	0.556
1.50	2.10	1.40	3.58	0.540
1.50	2.20	1.50	3.56	0.537

(c) RAP-MUSIC

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
1.50	2.20	1.40	11.94	1.0
1.50	2.20	1.50	4.67	0.391
1.50	2.20	1.30	3.76	0.315
1.60	2.10	1.30	3.57	0.299
1.60	2.20	1.40	3.04	0.254

表 3.5: セッション 3 (ギター音) の推定結果

(a) MUSIC

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
1.50	2.20	1.40	2.60	1.0
1.50	2.20	1.30	1.79	0.689
1.50	2.20	1.50	1.74	0.671
1.50	2.30	1.40	1.65	0.634
1.50	2.30	1.50	1.65	0.634

(b) 空間平均化法

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
1.50	2.20	1.40	2.43	1.0
1.50	2.30	1.40	1.68	0.689
1.50	2.30	1.50	1.62	0.664
1.50	2.20	1.30	1.58	0.650
1.50	2.10	1.40	1.56	0.640

(c) RAP-MUSIC

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
1.50	2.20	1.40	4.18	1.0
1.50	2.20	1.30	1.82	0.434
1.50	2.20	1.50	1.77	0.423
1.50	2.30	1.40	1.38	0.330
1.50	2.30	1.50	1.36	0.325

表 3.6: セッション 4 (音源 1 : ホワイトノイズ & 音源 2 : 女性音声) の推定結果

(a1) MUSIC(音源 1)

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
1.50	2.20	1.40	10.49	1.0
1.50	2.20	1.50	6.60	0.629
1.60	2.20	1.40	6.34	0.604
1.60	2.10	1.30	6.34	0.604
1.50	2.20	1.30	6.05	0.577

(a2) MUSIC(音源 2)

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
3.10	1.65	1.70	6.92	1.0
3.10	1.65	1.80	5.96	0.861
3.10	1.65	1.60	5.60	0.809
3.10	1.55	1.70	5.55	0.802
3.20	1.65	1.70	5.51	0.796

(b1) 空間平均化法 (音源 1)

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
1.50	2.20	1.40	9.41	1.0
1.50	2.20	1.50	6.45	0.686
1.50	2.30	1.50	6.42	0.682
1.60	2.10	1.30	6.38	0.678
1.60	2.20	1.40	6.18	0.657

(b2) 空間平均化法 (音源 2)

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
3.10	1.65	1.70	6.73	1.0
3.10	1.55	1.70	5.33	0.793
3.10	1.65	1.80	5.16	0.768
3.10	1.65	1.60	5.04	0.749
3.10	1.75	1.80	5.04	0.748

(c1) RAP-USIC(音源 1)

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
1.50	2.20	1.40	13.49	1.0
1.50	2.20	1.50	6.79	0.503
1.60	2.10	1.30	5.57	0.413
1.60	2.20	1.40	5.04	0.374
1.50	2.20	1.30	4.67	0.346

(c2) RAP-MUSIC(音源 2)

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
3.10	1.65	1.70	8.05	1.0
3.10	1.65	1.80	5.75	0.715
3.10	1.65	1.60	4.29	0.533
3.10	1.55	1.70	4.01	0.499
3.10	1.55	1.60	4.01	0.498

表 3.7: セッション 5 (音源 1 : 男性音声 & 音源 2 : 女性音声) の推定結果

(a1) MUSIC(音源 1)

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
1.50	2.20	1.40	9.98	1.0
1.50	2.20	1.50	4.60	0.461
1.50	2.20	1.30	4.48	0.449
1.60	2.10	1.30	4.39	0.440
1.60	2.20	1.40	4.15	0.415

(a2) MUSIC(音源 2)

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
3.10	1.65	1.70	3.29	1.0
2.90	1.75	1.50	2.98	0.905
3.30	1.65	1.70	2.96	0.899
2.90	1.65	1.80	2.96	0.899
3.30	1.65	1.60	2.95	0.897

(b1) 空間平均化法 (音源 1)

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
1.50	2.20	1.40	9.11	1.0
1.60	2.10	1.30	5.03	0.552
1.50	2.30	1.50	4.47	0.491
1.50	2.10	1.40	4.35	0.478
1.50	2.30	1.40	4.26	0.467

(b2) 空間平均化法 (音源 2)

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
3.10	1.65	1.70	2.92	1.0
3.10	1.45	1.80	2.72	0.932
2.90	1.45	1.90	2.63	0.902
2.90	1.45	1.80	2.60	0.891
3.20	1.45	1.90	2.59	0.889

(c1) RAP-USIC(音源 1)

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
1.50	2.20	1.40	16.53	1.0
1.50	2.20	1.50	5.80	0.351
1.60	2.10	1.30	4.76	0.288
1.50	2.20	1.30	4.42	0.267
1.60	2.20	1.40	3.66	0.222

(c2) RAP-MUSIC(音源 2)

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
3.10	1.65	1.70	2.70	1.0
2.90	1.75	1.50	1.87	0.695
2.90	1.65	1.50	1.78	0.661
3.00	1.65	1.50	1.77	0.658
3.00	1.55	1.50	1.76	0.654

表 3.3 から表 3.7 に示す結果より，MUSIC，空間平均化法，RAP-MUSIC の 3 方法共に反射音の存在する環境においても正しく音源位置を推定可能であることがわかる．

これまでは壁面と天井の 3 面を用いた検討であったが，常に多面を使うことができるとは限らない．天井のマイクロホン 1 面のみを用いた場合のセッション 4 の結果を表 3.8 に示す．マイクロホン配置は図 3.13 であり「・」が実チャンネル (45 チャンネル)「○」が仮想チャンネル (32 チャンネル) を表す．

セッション 4 の結果を壁面 3 面を使用した場合，天井 1 面のみを使用した場合それぞれを図 3.14，3.15 に示す．横軸は  $k$  の大きさ順位を表し，縦軸が  $k$  の値を表す．白色の棒が音源の存在しない位置での  $k$  の値であり，灰色の棒が正解位置での  $k$  の値である．

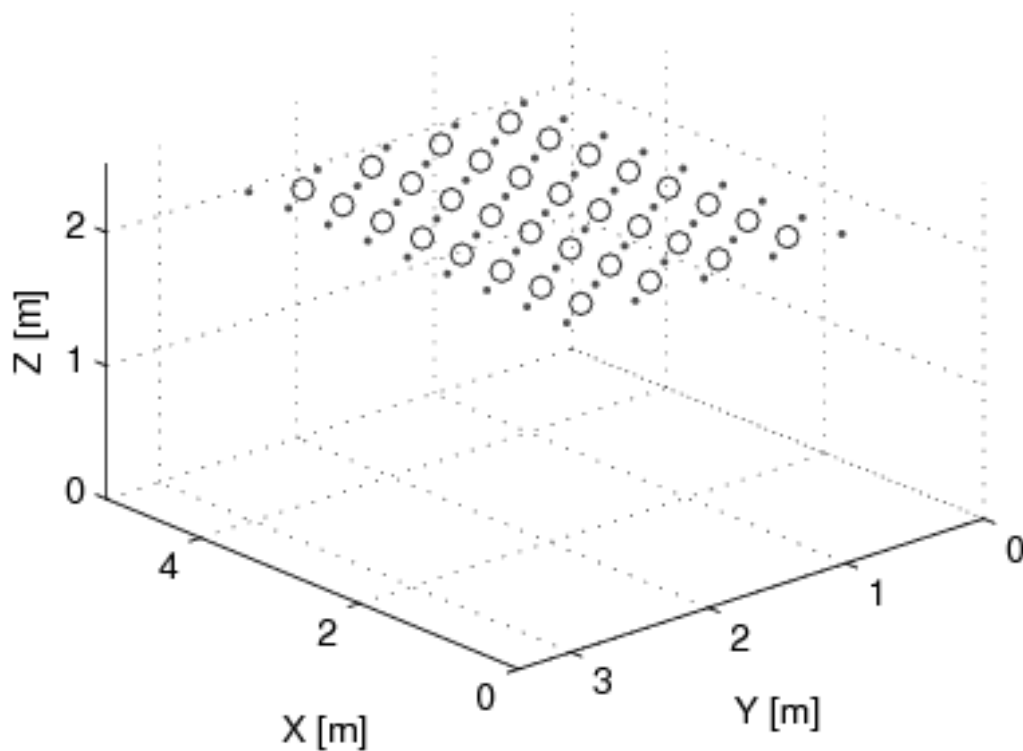


図 3.13: 実験に使用するマイクロホン (1 面) 「・」: 実チャンネル「○」: 仮想チャンネル

表 3.8: マイクロホン配置 1 面におけるセッション 4 (音源 1: ホワイトノイズ & 音源 2: 女性音声) の推定結果

(a1) MUSIC(音源 1)

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
1.50	2.20	1.30	9.35	1.137
1.50	2.20	1.40	8.22	1.0
1.50	2.20	1.20	7.44	0.905
1.60	2.20	1.40	6.38	0.776
1.60	2.20	1.50	6.12	0.745

(a2) MUSIC(音源 2)

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
1.50	2.20	1.30	9.35	1.137
1.50	2.20	1.40	8.22	1.0
1.50	2.20	1.20	7.44	0.905
1.60	2.20	1.40	6.38	0.776
1.60	2.20	1.50	6.12	0.745

(b1) 空間平均化法 (音源 1)

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
1.50	2.20	1.30	12.71	1.142
1.50	2.20	1.40	11.12	1.0
1.50	2.20	1.20	8.27	0.743
1.50	2.20	1.50	6.55	0.589
1.60	2.20	1.40	6.25	0.562

(b2) 空間平均化法 (音源 2)

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
3.10	1.65	1.70	10.09	1.0
3.10	1.65	1.60	6.04	0.599
3.10	1.65	1.80	5.85	0.580
3.10	1.65	1.90	4.51	0.447
3.10	1.65	1.50	4.13	0.409

(c1) RAP-USIC(音源 1)

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
1.50	2.20	1.40	13.69	1.0
1.50	2.20	1.30	13.37	0.976
1.50	2.20	1.20	9.15	0.668
1.50	2.20	1.50	8.98	0.656
1.60	2.20	1.40	6.44	0.471

(c2) RAP-MUSIC(音源 2)

$x$ (m)	$y$ (m)	$z$ (m)	$P$	$k$
3.10	1.65	1.70	12.98	1.0
3.10	1.65	1.60	8.60	0.662
3.10	1.65	1.80	7.14	0.550
3.10	1.65	1.50	6.58	0.507
3.10	1.65	1.90	4.96	0.382



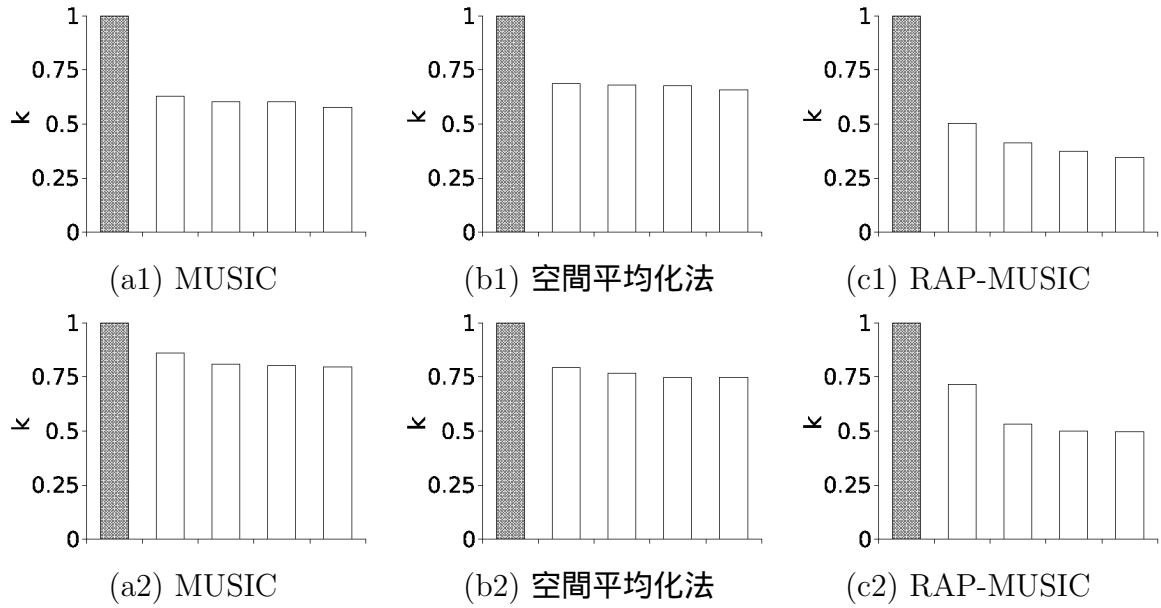


図 3.14: マイクロホン配置 3 面における推定結果 白色棒：音源位置ではない位置における  $k$  の値，灰色棒：音源位置における  $k$  の値

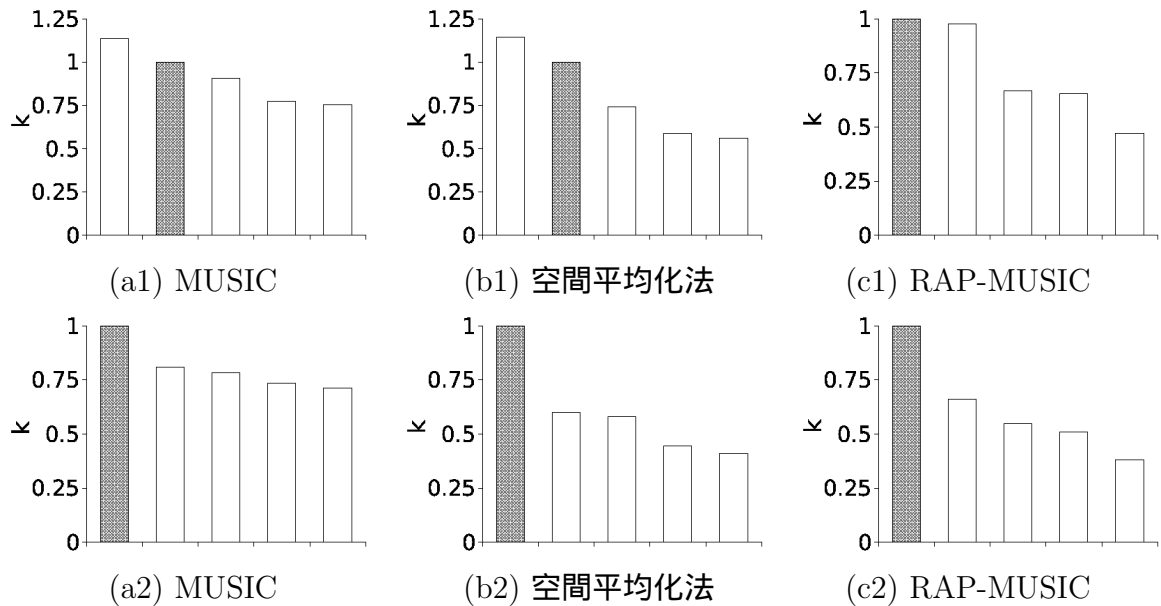


図 3.15: マイクロホン配置 1 面における推定結果 白色棒：音源位置ではない位置における  $k$  の値，灰色棒：音源位置における  $k$  の値

### 3.5.6 考察

表 3.3～表 3.7 や図 3.14 より，壁面 3 面のアレイを用いた場合は，MUSIC，空間平均化法，RAP-MUSIC のどの方法も音源の正解位置において  $P$  が一番大きな値を取り，正しい音源位置を推定していることがわかる．しかし，その中でも RAP-MUSIC はその他の探査点においても他の方法よりも  $k$  の値が小さく，音源位置のみでピークを形成していることが確認できる．さらに，天井 1 面を用いた場合，MUSIC や空間平均化法が誤った位置を推定してしまうのに対し，RAP-MUSIC のみ正しい位置を推定可能である．よって，RAP-MUSIC は他の方法よりも反射音に頑健な音源位置推定方法であると言える．

図 3.15 の (a1)，(b1)，(c1) を比較すると，アレイが 1 面の場合，従来法である MUSIC や空間平均化法では誤った位置に推定位置のピークを形成しているため，正しく音源位置を推定できていないことがわかる．マイクロホン配置が 1 面の場合，面に対して垂直方向のステアリングベクトルの変化量が小さいため，垂直方向の推定精度は高くない．よって，反射音の影響により，垂直方向に対して誤った位置を推定しまうためである．それに対して RAP-MUSIC のみ，全ての場合で正しい音源位置を推定できていることが確認できる．

## 3.6 まとめ

本章においては，実環境での音源位置において推定精度の低下につながる反射音の影響について検討し，反射音に頑健な音源位置推定の開発を行った．高精度な音源位置推定のためには，マイクロホン間隔に依存しない方法が望ましい．また，遅延和に基づく方法などの空間的なフーリエ変化に基づく方法の場合，高周波数での推定ではサイドローブが発生するという問題がある．そこで，受信信号の空間相関行列の固有空間構造を利用した MUSIC を用いるのが有益であることを示した．その上で，MUSIC による音源位置推定は反射音のある環境において精度が低下することを明らかにした．そこで，その対処方法として，空間平均化法により，MUSIC 法に用いる際の相関行列を多重平均することで，コヒーレント波のある環境においても高精度に MUSIC 法を行うことができることを示した．しかし，本研究において対象としているのは音源の到来方向推定ではなく 3 次元位置推定であるため，この方法をそのまま適用することができない．そこで，相関行列を探査位置ごとに計算し直す方法を提案し，その理論的な妥当性について検討した．さらに，この方法を発展させた，遅延和の原理と MUSIC 法を組み合わせた RAP-MUSIC を提案し，高精度推定ができる理論的な妥当性について検討した．この提案方法の推定能力を検討するために，計算機シミュレーションお

よび実環境における実験を行い，MUSIC，空間平均化法，RAP-MUSIC の性能評価を行った．その結果，RAP-MUSIC は従来の MUSIC 法よりも反射音の存在する環境における音源位置推定精度が高いことを示した．

## 第4章 室内残響下における原音情報の 高精度抽出法の提案

### 4.1 はじめに

本章では，室内で收音された多チャンネルの收音データから高精度に音源の原音情報を抽出する信号処理法について検討を行う．

小節 1.6.3 でも述べたとおり，高精度に原音情報を抽出するために，マイクロホン間隔に依存して空間的折り返し歪みを生じる欠点を持つビームフォーミングに基づく方法ではなく，室内伝達特性の逆フィルタを用いた残響除去による方法を検討する．具体的には，観測信号のみから室内伝達特性の逆特性を推定し，残響を高精度に除去できるブラインド残響除去についての検討を行う．

原音情報に対する先見的知識を用いないブラインド残響除去では，出力結果が白色化してしまうという問題がある．この問題に対して，線形予測に基づき観測信号から入力信号の平均を推定し，出力が白色化されず，高精度な残響除去が可能である LIME アルゴリズム (Linear-predictive Multi-input Equalization) が提案されている [70, 71] ．

一方，LIME を含むブラインド残響除去に関する研究の多くは，その後段に音声認識を想定していることが多く，検討されているサンプリング周波数は 8 kHz ~ 16 kHz としていることが多い．しかし，本研究では，楽器音や物音などの様々な音響事象での残響除去を想定しているため，ブラインド残響除去のさらなる高帯域化が必要である．

音声や楽器，物音などといった一般の音は，サンプリング周波数を高くするにつれて周波数軸方向でエネルギーの高い部分と低い部分とがまばらに存在する．このような観測信号の場合，従来の LIME アルゴリズムでは残響除去精度が劣化するという問題がある．しかし，エネルギーは低いとはいえ，肉声感や場の雰囲気などを伝える際に重要な成分であることは間違いない．そこで本章では，高帯域での高精度な残響除去アルゴリズムの開発を行う．

4.2 節において，本研究で対象としたブラインド残響除去の基本原則である MINT について概要する．4.3 節では，観測信号のみから系の伝達関数 (=インパルス応答) を推定するブラインドシステム同定を用いた残響除去に関する先行研究，およびその問題点について述べる．4.4 節では，系の伝達関数の逆特性を入力信号から直接推定するブ

ラインド残響除去の原理と，本章での検討課題となる LIME アルゴリズムについて述べる．4.5 節では LIME アルゴリズムの問題点を指摘し，その問題点を解決するため Pre-Whitening フィルタを用いた方法を提案し，シミュレーションによる性能評価を行う．4.6 節で本章のまとめを述べる．

## 4.2 MINT による完全な逆フィルタと残響除去

### 4.2.1 非最小位相系フィルタ

室内の残響を完全に取り除くためには，音源と受音点との室内伝達関数 (=インパルス応答) の逆特性 (=逆フィルタ) を受音信号に畳み込むことによって実現できる．逆フィルタとは，あるシステムの伝達関数を完全にキャンセルすることのできるフィルタである．

音源と受音点間のインパルス応答の  $z$  変換を  $H(z)$  とし，逆フィルタを  $G(z)$  とすると，逆フィルタ  $G(z)$  とは，

$$H(z)G(z) = 1 \quad (4.1)$$

を満たすフィルタのことである．しかし，室内インパルス応答は応答長 (=タップ長) が非常に長いため，非最小位相系 (nonminimum-phase) となる場合が多い [72]．

最小位相系 (minimum-phase system) とは，伝達関数の全ての零点が，複素  $z$  平面上の単位円内に含まれているものをいう．ちなみに  $G(z)$  の零点とは  $G(z) = 0$  となる  $z$  の値である．非最小位相系の逆フィルタは，因果性を満たした不安定なものか，または，安定ではあるが非因果なものとなる．非最小位相の零点が逆フィルタの安定性に及ぼす影響は，定性的には，以下のように解釈することができる．まず，伝達関数  $G(z)$  の零点  $z_q$  が  $z$  平面上の単位円周上にあつて， $z_q = e^{j\omega_q}$  と表される場合には，周波数  $\omega_q$  の信号に対する  $G(z)$  の応答は零となる．このとき， $G(z)$  の逆特性は，応答が零で消えてしまった周波数成分を回復するために，その周波数  $\omega_q$  で無限大のゲインをもつことになり，不安定なフィルタとなる．次に，零点が単位円外にある場合は，対応する周波数成分を中心として信号が遅延することを意味している．このような遅延した信号を回復するには，「時間を進ませる」という現実的には不可能な操作を行う必要がある．その結果，逆フィルタは非因果性となるか，または因果性を満たすように実現しようとしたときには不安定なものになってしまう [50]．

したがって，1 入力 1 出力の室内音響システムの場合，その逆フィルタは近似解でしか求めることができない [74–77]．

#### 4.2.2 MINT

これに対し，Multiple-input/output INverse Theorem(MINT) を用いることにより，伝達系が非最小位相系であっても音源数よりも多い複数のマイクロホンを用いて安定な逆フィルタを逆行列演算によって計算できる [73]．MINT は，多入力多出力理論であり，複数音源でも対応可能であるが，ここでは 1 入力  $P$  出力の音響システムの場合について述べる．

音源とマイクロホン  $i$  間のインパルス応答を  $h_i(n)$  とし，タップ長を  $M$  として，時間軸上の畳み込みを考える．インパルス応答行列  $\mathbf{H}$  を以下のように定義する．

$$\mathbf{H} = [\mathbf{H}_1, \dots, \mathbf{H}_P] \quad (4.2)$$

$$\mathbf{H}_i = \begin{pmatrix} h_i(0) & 0 & \dots & 0 \\ h_i(1) & h_i(0) & \ddots & \vdots \\ \vdots & & \ddots & \\ h_i(M-1) & & & 0 \\ 0 & h_i(M-1) & & h_i(0) \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & h_i(M-1) \end{pmatrix} \quad (4.3)$$

逆フィルタ行列  $\mathbf{G}$  を以下のように定義する．

$$\mathbf{G} = [\mathbf{g}_1^T, \dots, \mathbf{g}_P^T]^T \quad (4.4)$$

ここで，

$$\mathbf{g}_i = [g_i(0), \dots, g_i(L-1)]^T \quad (4.5)$$

である．逆フィルタ長  $L$  が，

$$L = \frac{M-1}{P-1} \quad (4.6)$$

を満たす時，行列  $\mathbf{H}$  は正方行列となる．室内インパルス応答の伝達関数の零点が重なっておらず，式 (4.6) を満たすとき，逆フィルタ  $\mathbf{G}$  は，

$$\mathbf{G} = \mathbf{H}^{-1} \mathbf{B} \quad (4.7)$$

$$\mathbf{B} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (4.8)$$

となり， $\mathbf{H}$  の逆行列演算から算出可能となる．この結果，

$$\sum_{i=1}^P H_i(z) G_i(z) = 1 \quad (4.9)$$

となる．よって，多入力多出力系を用いることにより，完全な逆フィルタの算出が可能である．

MINTにより算出した逆フィルタを用いることにより室内残響を完全に取り除くことができることを示す．

1入力  $P$  出力の音響システムを図 4.1 に示す．

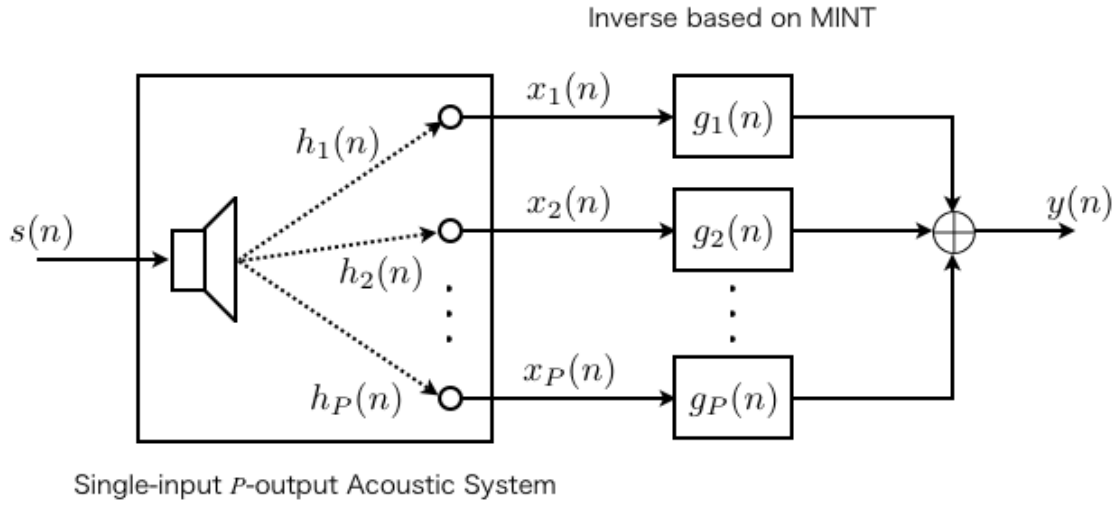


図 4.1: 1 入力  $P$  出力音響システムに対する MINT による残響除去処理のブロック図 ( $s(n)$ : 原音信号,  $h_i(n)$ : 室内伝達関数,  $x_i(n)$ : 観測信号,  $g_i(n)$ : MINT 型逆フィルタ,  $y(n)$ : 回復信号)

音源信号を  $s(n)$ ，音源とマイクロホン  $i$  間のインパルス応答を  $h_i(n)$ ，マイクロホン  $i$  での入力信号を  $x_i(n)$  とし，それぞれの  $z$  変換を  $S(z)$ ， $H_i(z)$ ， $X_i(z)$  とすると，

$$X_i(z) = S(z)H_i(z) \quad (4.10)$$

となる．よって，式 (4.9) より，

$$\sum_{i=1}^P X_i(z)G_i(z) = \sum_{i=1}^P S(z)H_i(z)G_i(z) = S(z) \quad (4.11)$$

となり，MINT で算出した逆フィルタによって音源信号を完全に回復できることがわかる．

しかし，この方法は室内インパルスが測定されたときにのみ有効であるため，逆フィルタを有効に利用できる環境に制限がある．このため，マイクロホンでの観測信号のみから室内インパルスを推定するブラインドシステム同定や観測信号から直接逆フィルタを推定するブラインド残響除去が検討されている．

## 4.3 ブラインド多チャネルシステム同定によるインパルス応答の推定

### 4.3.1 ブラインド多チャネルシステム同定

ブラインドシステム同定とは，音源位置，原音情報などの先見的知識を必要とせず，観測信号のみから系のインパルス応答を推定する方法であり，Sato によって提案された [78] ．

ブラインドシステム同定には，これまでも様々な方法が検討されている．文献 [78] や [79] に代表される Constant Modulus Algorithm (CMA) に基づく方法や高次統計量に基づく方法 [80] は評価関数が多峰性であるため，局所解 (Local Minimum) に陥りやすいといった欠点を持つ．そのため，多くは観測信号の 2 次モーメントを用いた 2 次統計量に基づく方法が検討されている [81–84] ．

2 次統計量に基づく方法の原理は以下の通りである．背景雑音がない状態を仮定すると，チャンネル  $i$  における観測信号  $x_i(n)$  は，原音信号を  $s(n)$ ，音源位置からマイクロホン  $i$  までのインパルス応答を  $h_i$  とすると， $s(n)$  と  $h_i$  の線形畳み込みによって

$$x_i(n) = s(n) * h_i(n) \quad (4.12)$$

と記述できる．式 (4.12) の両辺に  $i$  チャンネルとは異なるチャンネル  $j$  のインパルス応答を  $h_j(n)$  を畳み込むと，

$$x_i(n) * h_j(n) = s(n) * h_i(n) * h_j(n) = s(n) * h_j(n) * h_i(n) = x_j(n) * h_i(n) \quad (4.13)$$

となる．つまり，式 (4.13) の関係から，次項に紹介する方法などを用いて， $h_i(n)$  や  $h_j(n)$  を推定することが可能となる．

ブラインドシステム同定はもともとインパルス応答が短い無線通信などの分野で多くの検討が行われてきたが，近年ではインパルス応答が長い音響システムの分野でも非常に注目されている方法である．以下では，2 次統計量を用いた室内インパルス推定方法の中でも代表的な Least Mean Square (LMS) [85] を用いて適応的にインパルスを推定する方法と部分空間 (サブスペース) に基づく方法について述べる．

### 4.3.2 LMS アルゴリズムに基づく方法

本項では，時間軸領域における Multi-Channel LMS (MCLMS) [86] を例にして LMS アルゴリズムに基づく方法について述べる．



ここで用いる SIMO(Single-Input Multiple-Output) モデルも 4.2 節と同様，図 4.1 のように表す． $i$  チャンネルの出力信号  $x_i(n)$  は音源信号  $s(n)$  と音源位置から出力位置までの伝達関数 (=インパルス応答) の線形畳み込みで表される．

$$x_i(n) = h_i * s(n), \quad i = 1, 2, \dots, P \quad (4.14)$$

ここで， $P$  はマイクロホン数である．ベクトル形式では，

$$\mathbf{x}_i(n) = \mathbf{H}_i \cdot \mathbf{s}(n) \quad (4.15)$$

ここで，一番長いインパルス応答の長さを  $M$  とすると，

$$\mathbf{x}_i(n) = [x_i(n) \quad x_i(n-1) \quad \dots \quad x_i(n-M+1)]^T \quad (4.16)$$

$$\mathbf{H}_i = \begin{bmatrix} h_{i,0} & h_{i,1} & \dots & h_{i,M-1} & 0 & \dots & 0 \\ 0 & h_{i,0} & \dots & h_{i,M-2} & h_{i,M-1} & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & h_{i,0} & h_{i,1} & \dots & h_{i,M-1} \end{bmatrix} \quad (4.17)$$

$$\mathbf{s}(n) = [s(n) \quad s(n-1) \quad \dots \quad s(n-M+1) \quad \dots \quad s(n-2M+2)]^T \quad (4.18)$$

となる． $i$  チャンネルのインパルス応答は

$$\mathbf{h}_i = [h_{i,0} \quad h_{i,1} \quad \dots \quad h_{i,M-1}]^T \quad (4.19)$$

である．

ここで，式 (4.14) から，

$$x_i * h_j = s * h_i * h_j = x_j * h_i, \quad i, j = 1, 2, \dots, P \quad (4.20)$$

という関係が得られる．時刻  $n$  において，

$$\mathbf{x}_i^T(n) \mathbf{h}_j = \mathbf{x}_j^T(n) \mathbf{h}_i, \quad i, j = 1, 2, \dots, P \quad (4.21)$$

という関係が得られ，両辺に左側から  $\mathbf{x}_i$  を掛けると，

$$\mathbf{R}_{x_i x_i} \mathbf{h}_i = \mathbf{R}_{x_i x_j} \mathbf{h}_j, \quad i, j = 1, 2, \dots, P \quad (4.22)$$

となる．ここで， $\mathbf{R}_{x_i x_j} = E\{\mathbf{x}_i(n) \mathbf{x}_j^T(n)\}$  である．よって， $P(P-1)$  個の式から，

$$\sum_{i=1, i \neq j}^P \mathbf{R}_{x_i x_i} \mathbf{h}_i = \sum_{i=1, i \neq j}^P \mathbf{R}_{x_i x_j} \mathbf{h}_j, \quad i, j = 1, 2, \dots, P \quad (4.23)$$

となる．さらに，

$$\mathbf{R}\mathbf{h} = \mathbf{0} \quad (4.24)$$

という関係が得られる．ここで，

$$\mathbf{R} = \begin{bmatrix} \sum_{i \neq 1} \mathbf{R}_{x_i x_i} & -\mathbf{R}_{x_2 x_1} & \cdots & -\mathbf{R}_{x_M x_1} \\ -\mathbf{R}_{x_1 x_2} & \sum_{i \neq 2} \mathbf{R}_{x_i x_i} & \cdots & -\mathbf{R}_{x_M x_2} \\ \vdots & \vdots & \ddots & \vdots \\ -\mathbf{R}_{x_1 x_P} & -\mathbf{R}_{x_2 x_P} & \cdots & \sum_{i \neq P} \mathbf{R}_{x_i x_i} \end{bmatrix} \quad (4.25)$$

$$\mathbf{h} = [\mathbf{h}_1^T \quad \mathbf{h}_2^T \quad \cdots \quad \mathbf{h}_P^T]^T \quad (4.26)$$

である．

適応的に推定するインパルス応答を

$$\hat{\mathbf{h}} = [\hat{\mathbf{h}}_1^T \quad \hat{\mathbf{h}}_2^T \quad \cdots \quad \hat{\mathbf{h}}_P^T]^T \quad (4.27)$$

とすると，MCLMSにおいて，式(4.21)から，

$$\epsilon_{ij}(n) = \begin{cases} \mathbf{x}_i^T(n)\hat{\mathbf{h}}_j - \mathbf{x}_j^T(n)\hat{\mathbf{h}}_i, & i \neq j, i, j = 1, 2, \dots, P \\ 0, & i = j, i, j = 1, 2, \dots, P \end{cases} \quad (4.28)$$

をエラー信号とすると，コスト関数は，

$$J(n) = \sum_{i=1}^{P-1} \sum_{j=i+1}^P \epsilon_{ij}^2(n) \quad (4.29)$$

となる．適応アルゴリズムの拘束条件は，

$$\hat{\mathbf{h}} = \arg \min \mathbf{E}\{J(n)\} \quad (4.30)$$

であり，MCLMSは以下のようにインパルス応答 $\hat{\mathbf{h}}$ を更新する．

$$\hat{\mathbf{h}}(n+1) = \frac{\hat{\mathbf{h}}(n) - 2\mu[\tilde{\mathbf{R}}(n)\hat{\mathbf{h}}(n) - J(n)\hat{\mathbf{h}}]}{\|\hat{\mathbf{h}}(n) - 2\mu[\tilde{\mathbf{R}}(n)\hat{\mathbf{h}}(n) - J(n)\hat{\mathbf{h}}]\|} \quad (4.31)$$

ここで， $\mu$ はステップサイズ，

$$\tilde{\mathbf{R}} = \begin{bmatrix} \sum_{i \neq 1} \tilde{\mathbf{R}}_{x_i x_i} & -\tilde{\mathbf{R}}_{x_2 x_1} & \cdots & -\tilde{\mathbf{R}}_{x_P x_1} \\ -\tilde{\mathbf{R}}_{x_1 x_2} & \sum_{i \neq 2} \tilde{\mathbf{R}}_{x_i x_i} & \cdots & -\tilde{\mathbf{R}}_{x_P x_2} \\ \vdots & \vdots & \ddots & \vdots \\ -\tilde{\mathbf{R}}_{x_1 x_P} & -\tilde{\mathbf{R}}_{x_2 x_P} & \cdots & \sum_{i \neq P} \tilde{\mathbf{R}}_{x_i x_i} \end{bmatrix} \quad (4.32)$$

$$\mathbf{R}_{x_i x_j} = \mathbf{E}\{\mathbf{x}_i(n)\mathbf{x}_j^T(n)\}, \quad i, j = 1, 2, \dots, P \quad (4.33)$$

である．

MCLMS とは，このようにして適応的に  $\hat{\mathbf{h}}$  を更新させるアルゴリズムである．時間領域での MCLMS のステップサイズを最適にした方法も提案されている [87]．また，MCLMS を  $2L$  フレームずつ切り出し，円状畳み込みとオーバーラップによって  $L$  フレームずつインパルス応答  $\hat{\mathbf{h}}$  を周波数軸領域で求めるの Multichannel Frequency-domain LMS(MCFLMS) も提案されている [88–91]．

これらの方法により，系のインパルス応答を適応的に推定することができる．しかし，インパルス応答長が長い場合，収束速度が極めて遅くなるといった欠点を持つ．

### 4.3.3 部分空間法を用いた伝達関数の推定

観測信号のみから室内インパルス応答を推定する方法としては，LMS による方法以外に，部分空間法 (サブスペース法) を用いた方法がある [92–94]．ここでは，2 チャネルの入力信号から得られる入力相関行列からインパルス応答を推定する方法 [93] について述べる．

入力信号を  $s(n)$ ，音源からマイクロホン 1, 2 までの室内伝達関数をそれぞれ  $h_1(n)$ ， $h_2(n)$ ，マイクロホン 1, 2 の入力信号をそれぞれ  $x_1(n)$ ， $x_2(n)$  とする．図 4.2 のように  $x_1(n)$ ， $x_2(n)$  がそれぞれフィルタ  $\hat{h}_2(n, i)$ ， $\hat{h}_1(n, i)$  を通ったあと，差分信号  $e(n)$  が得られる． $h_1(n)$ ， $h_2(n)$  は次数  $d$  の FIR フィルタでモデル化できるとし，互いに共通零点を持たないと仮定する．このとき， $e(n, i)$  は

$$\begin{aligned} e(n, i) &= x_1(n) * \hat{h}_2(n, i) - x_2(n) * \hat{h}_1(n, i) \\ &= s(n) * h_1(n) * \hat{h}_2(n, i) - s(n) * h_2(n) * \hat{h}_1(n, i) \\ &= s(n) * \{h_1(n) * \hat{h}_2(n, i) - h_2(n) * \hat{h}_1(n, i)\} \end{aligned} \quad (4.34)$$

となる．式 (4.34) より， $i = d$  が成立し，かつ，すべての  $n$  に対して  $e(n, i) = 0$  であるならば，

$$\begin{aligned} \hat{h}_1(n, i) &= \alpha h_1(n) \\ \hat{h}_2(n, i) &= \alpha h_2(n) \end{aligned} \quad (4.35)$$

が成り立つ．ここで， $\alpha$  は任意定数である．逆に，式 (4.35) が成り立てば， $e(n, i) = 0$  となる．

$e(n, i)$  は計算誤差，測定誤差などにより正確に 0 とはならないので， $e(n, i)$  の二乗平均誤差を最小にするような  $\hat{h}_1(n, i)$ ， $\hat{h}_2(n, i)$  を計算する． $e(n, i)$  の二乗平均誤差は

$$\begin{aligned} E\{e^2(n, i)\} &= E\{\hat{\mathbf{h}}^T(i) \mathbf{x}(i) \mathbf{x}^T(i) \hat{\mathbf{h}}(i)\} \\ &= \hat{\mathbf{h}}^T(i) \mathbf{R}(i) \hat{\mathbf{h}}(i) \end{aligned} \quad (4.36)$$

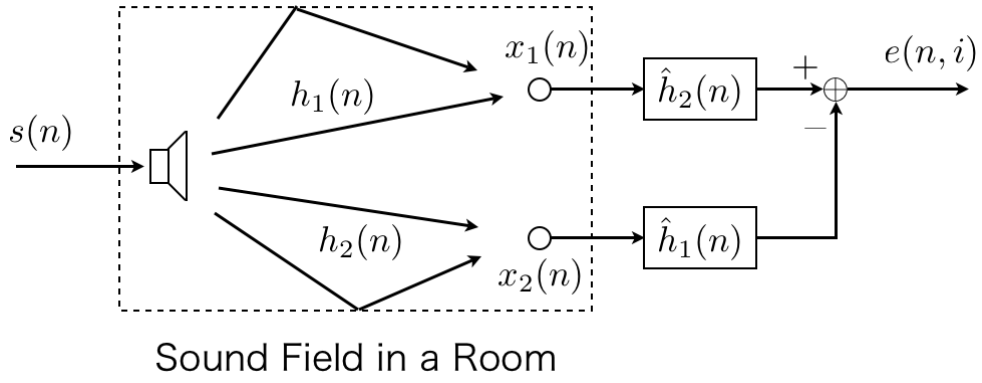


図 4.2: 推定した室内伝達関数による誤差信号  $e(n, i)$  の算出 ( $s(n)$ : 原音信号,  $h_i(n)$ : 室内伝達関数,  $x_i(n)$ : 観測信号,  $\hat{h}_i(n)$ : 推定した室内伝達関数,  $e(n, i)$ : 誤差信号)

ここで,  $\hat{\mathbf{h}}$  はフィルタ係数ベクトル:

$$\hat{\mathbf{h}}(i) = \begin{pmatrix} \hat{h}_2(0, i) \\ \hat{h}_2(1, i) \\ \vdots \\ \hat{h}_2(i, i) \\ -\hat{h}_1(0, i) \\ -\hat{h}_1(1, i) \\ \vdots \\ -\hat{h}_1(i, i) \end{pmatrix} \quad (4.37)$$

$\mathbf{x}(i)$  は入力信号ベクトル:

$$\mathbf{x}(i) = \begin{pmatrix} x_1(n) \\ x_1(n-1) \\ \vdots \\ x_1(n-i) \\ x_2(n) \\ x_2(n-1) \\ \vdots \\ x_2(n-i) \end{pmatrix} \quad (4.38)$$

$\mathbf{R}$  は入力相関行列:

$$\mathbf{R} = E\{\mathbf{x}(i)\mathbf{x}^T(i)\} \quad (4.39)$$

であり， $E\{\}$  は期待値をあらわす．ここで， $e(n, i)$  が 0 となるとき，式 (4.34) は，

$$\mathbf{R} \begin{pmatrix} \hat{h}_2(0, i) \\ \hat{h}_2(1, i) \\ \vdots \\ \hat{h}_2(i, i) \\ -\hat{h}_1(0, i) \\ -\hat{h}_1(1, i) \\ \vdots \\ -\hat{h}_1(i, i) \end{pmatrix} = 0 \quad (4.40)$$

となる．式 (4.40) の意味するところは， $e(n, i)$  の二乗平均誤差を  $|\mathbf{h}(i)|$  一定で最小化するフィルタ係数  $\mathbf{h}(i)$  は，入力相関行列  $\mathbf{R}(i)$  の最小固有値に対する最小固有ベクトルとして計算される．

伝達関数  $h_1(n)$ ， $h_2(n)$  の次数は未知であり，フィルタ次数が適当でなければ式 (4.35) は成り立たない．そこで，フィルタ次数を決定するために図 4.3 に示すような二層フィルタ構造から導かれる評価関数を導入する．

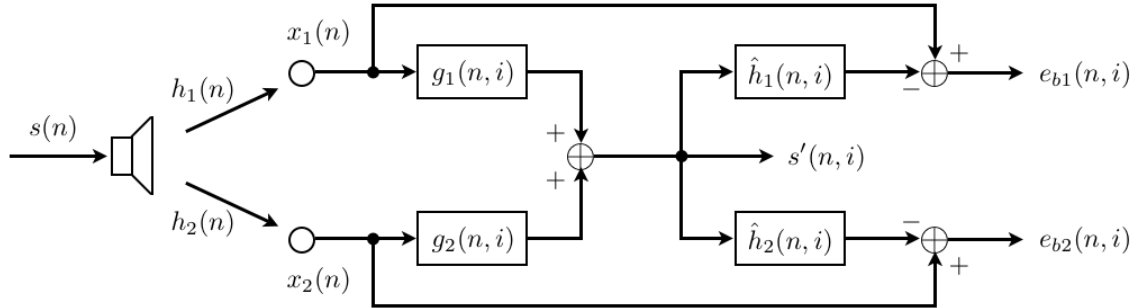


図 4.3: 部分空間法における最適な次数  $i$  の探索

図 4.3 において， $g_1(n, i)$ ， $g_2(n, i)$  は  $\hat{h}_2(n, i)$ ， $\hat{h}_1(n, i)$  の MINT 型逆フィルタ [73] であり，式 (4.11) の  $P = 2$  である次式を解くことによって計算される．

$$G_1(z, i)H_1(z, i) + G_2(z, i)H_2(z, i) = 1 \quad (4.41)$$

ここで， $G_1(z, i)$ ， $H_1(z, i)$ ， $G_2(z, i)$ ， $H_2(z, i)$  はそれぞれ  $g_1(n, i)$ ， $g_2(n, i)$ ， $\hat{h}_2(n, i)$ ， $\hat{h}_1(n, i)$  の  $z$  変換である．次数を  $i$  として得られたときの  $g_1(n, i)$ ， $g_2(n, i)$  を用いて，音源信号  $s(n)$  の残響回復を行った信号を  $s'(n, i)$  とすると，

$$s'(n, i) = x_1(n) * g_1(n, i) - x_2(n) * g_2(n, i) \quad (4.42)$$

となる．このとき，次式の評価関数を採用する．

$$PE(i) = \frac{E\{e_{b1}^2(n, i)\}}{E\{x_1^2(n)\}} + \frac{E\{e_{b2}^2(n, i)\}}{E\{x_2^2(n)\}} \quad (4.43)$$

ここで，

$$\begin{aligned} e_{b1}^2(n, i) &= x_1(n) - s'(n, i) * \hat{h}_1(n, i) \\ e_{b2}^2(n, i) &= x_2(n) - s'(n, i) * \hat{h}_2(n, i) \end{aligned}$$

この評価関数は，各次数  $i$  に対して推定された  $\hat{h}_2(n, i)$ ， $\hat{h}_1(n, i)$  と  $s'$  を用いて，実際の残響信号  $x_1(n) = s(n) * h_1(n)$ ， $x_2(n) = s(n) * h_2(n)$  をどの程度よくあらわせるかを評価するものである．このとき， $PE(i) = 0$  つまり  $e_{b1}(n, i) = 0$ ， $e_{b2}(n, i) = 0$  ならば，またそのときに限り次式，

$$\begin{aligned} \hat{h}_1(n, i) &= \alpha h_1(n) \\ \hat{h}_2(n, i) &= \alpha h_2(n) \\ s'(n, i) &= \frac{1}{\alpha} s(n) \end{aligned}$$

が成り立つ．よって， $PE(i)$  を最小にする最適次数は，

$$I = \arg \min(PE(i)) \quad (4.44)$$

と表すことができ， $\hat{h}_1(n, i)$ ， $\hat{h}_2(n, i)$ ， $s'(n, i)$  を最終的な推定値とする．

#### 4.3.4 先行研究のまとめと問題点

このように，観測信号の 2 次モーメントを利用した 2 次統計量に基づく方法を用いることによって，観測信号のみからインパルス応答を推定することができる．しかし，これらの方法は数学的に解くことができるが，実環境でのインパルス応答が推定が非常に困難であるのが現状である．その一番の原因は，室内インパルスは非常に長い応答であることである．応答長はサンプリング周波数に依存するが，室内残響のインパルス応答は数千タップ，高サンプリングであれば数万タップといった長さとなる．そのため，LMS に基づく方法では推定するインパルス応答のタップ長が長いほど収束速度は遅くなり，数千タップの高精度なインパルス応答を適応的に求めるのは困難であるのが現状である．また，サブスペースを用いた方法で問題となるのは，4.3.3 項でも述べた通り，インパルス応答のタップ数を正確に把握する必要がある点である．数千タップの長さをもつインパルス応答に対して，文献 [93] のように総当たりにタップ長を変えて推定するのは現実的ではない．インパルス応答が長い問題については，音

響エコーキャンセラーでも用いられているフィルタバンク [95,96] を用いた試みが行われている。これは、完全再構成フィルタバンク [95] などを導入することによって、帯域ごとに間引かれた短いインパルス応答を推定する問題に帰着することができるためである。エコーキャンセラではフィルタバンク [97] を導入することにより精度よくインパルスを推定できる [98]。しかし、ブラインド多チャネルシステム同定の場合、系の入力信号も不明であるため、式 (4.35) の  $\alpha$  が帯域ごとに異なる値として計算される。そのため、 $\alpha$  を別途推定する必要があるが、その方法は未開拓である [93,99]。これは、独立成分分析 [100,101] に基づく周波数領域ブラインド音源分離 [102] における各周波数帯域ごとのスケールリング問題と同様の問題である。

これらの理由から、観測信号のみから室内インパルス応答を推定するのは困難であるのが現状である。

## 4.4 ブラインド 残響除去

### 4.4.1 Semi-blind MINT

そのような理由から、系のインパルス応答は求めるのではなく、観測信号のみから系のインパルスの逆フィルタを直接算出する Semi-blind-MINT が提案されている [103,104]。

$P$  チャネルの入力信号  $\mathbf{X} = [\mathbf{X}_1 \mathbf{X}_2 \cdots \mathbf{X}_P]$  に対して入力相関行列  $\mathbf{R}$  を、

$$\begin{aligned} \mathbf{R} &= \mathbf{E}\{\mathbf{X}^T \mathbf{X}\} \\ &= \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \cdots & \mathbf{R}_{1P} \\ \mathbf{R}_{21} & \mathbf{R}_{22} & \cdots & \mathbf{R}_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{R}_{P1} & \mathbf{R}_{P2} & \cdots & \mathbf{R}_{PP} \end{bmatrix} \end{aligned} \quad (4.45)$$

と定義する。ただし、

$$\begin{aligned} \mathbf{X}_j &= [x_j(n) x_j(n-1) \cdots x_j(n-(L-1))] \\ \mathbf{R}_{ij} &= \begin{bmatrix} r_{ij}(0) & r_{ij}(-1) & \cdots & r_{ij}(-L+1) \\ r_{ij}(1) & r_{ij}(0) & \cdots & r_{ij}(-L+2) \\ \vdots & \vdots & \ddots & \vdots \\ r_{ij}(L-1) & r_{ij}(L-2) & \cdots & r_{ij}(0) \end{bmatrix} \end{aligned}$$

である。 $r_{ij}(\tau)$  は  $i$  チャネルの観測信号  $x_i(n)$  と  $j$  チャネルの観測信号  $x_j(n)$  の相関関数を表し、

$$r_{ij}(\tau) = \mathbf{E}\{x_i(n)x_j(n+\tau)\}$$

である． $L$ は逆フィルタ長である．したがって  $\mathbf{R}$  は  $PL \times PL$  行列である．ここで入力信号  $\mathbf{X}$  は，音源信号  $\mathbf{S}$  と式 (4.3) の室内インパルス応答  $\mathbf{H}$  を用いて，

$$\mathbf{X} = \mathbf{S}\mathbf{H} \quad (4.46)$$

であり， $\mathbf{S} = [s(n)s(n-1)\cdots s(n-(M+L-2))]$  である．式 (4.45) を式 (4.46) を代入すると，

$$\mathbf{R} = \mathbf{E}\{\mathbf{H}^T \mathbf{S}\mathbf{H}\} = \mathbf{H}^T \mathbf{E}\{\mathbf{S}^T \mathbf{S}\} \mathbf{H} \quad (4.47)$$

となる．ここで，入力信号が白色の場合，

$$\mathbf{E}\{s(n)s(n+\tau)\} = \delta(\tau) \quad (4.48)$$

となり，

$$\mathbf{E}\{\mathbf{S}^T \mathbf{S}\} = \mathbf{I} \quad (4.49)$$

となる． $\delta(\tau)$  はデルタ関数， $\mathbf{I}$  は単位行列である．このとき，

$$\mathbf{R} = \mathbf{H}^T \mathbf{H} \quad (4.50)$$

となる．式 (4.7) と式 (4.50) より，

$$\mathbf{R}\mathbf{G} = \mathbf{H}^T \mathbf{H}\mathbf{H}^{-1} \mathbf{B} = \mathbf{H}^T \mathbf{B} \quad (4.51)$$

となる．ここで，音源がマイクロホン 1 ともっとも近いとすると，最初に到達する  $h_1(0)$  以外の  $h_i(0)$  はすべて 0 となる．このとき，式 (4.51) 右辺の  $\mathbf{H}^T$  の最左列は室内インパルス応答の先頭  $h_1(0)$  以外は 0 となり，

$$\mathbf{H}^T \mathbf{B} = h_1(0) \mathbf{B} \quad (4.52)$$

となり，音源からの音がマイクロホン 1 に最も最初に到達していることを示す．式 (4.51) と式 (4.52) から逆フィルタ  $\mathbf{G}$  は，

$$\mathbf{G} = h_1(0) \mathbf{R}^{-1} \mathbf{B} \quad (4.53)$$

と相関行列のみから計算できる．

Semi-blind-MINT などの逆フィルタを直接求める方法の特徴として，ブラインドシステム同定で問題となっていたインパルス応答の正確な長さがわからなくてもよいというメリットがある．Semi-blind-MINT では，逆フィルタ長  $L$  を式 (4.6) で定義される実際のインパルス応答長に相当する長さよりも長く見積もって計算しても逆フィル



タの精度は劣化しないことが確認されている [103,104] . また , タップ長が数千タップと長い場合においても音源が白色信号であれば良好な残響除去を実現できる . さらに , 雑音環境においてもロバストに動作することが報告されている [103,104] .

Semi-blind-MINT が完全にブラインド処理でない理由は , 式 (4.53) で示す通り , 音源と一番距離が近いマイクロホンのチャンネル番号が必要な点である . しかしこの点は , 第 3 章で検討した高精度な音源位置推定を用いることにより解決可能である .

しかし , Semi-blind-MINT では , 音源信号の白色性を仮定して逆フィルタ  $G$  を計算しているため , 音源が白色信号でない場合は残響除去精度が劣化するという問題がある . 音声など , 一般に存在する音源は白色信号ではなく , 音源特有の平均スペクトルを有する . したがって , 音声信号などの有色信号にそのまま Semi-blind-MINT を適用すると , 音源の平均スペクトルも除去され , 出力結果が白色化してしまう問題が生じる . 文献 [103,104] ではこの問題を解決するために , 観測信号  $X$  の平均スペクトルで最初に  $X$  を白色化している , しかし , 観測信号  $X$  の平均スペクトルには室内インパルス応答  $H$  のスペクトル成分も混入しているため , 完全な残響除去にはならないという問題がある .

#### 4.4.2 線形予測を用いた高精度ブラインド 残響除去

雑音の存在環境において , 単一音源環境におけるブラインド多チャンネル残響除去では , 以下の 3 点を解決することが望まれる .

- i) インパルス応答長が大きい ( ~ 数千タップ ) ことへの対応
- ii) インパルス応答のタップ長が不明であっても推定可能であること
- iii) 出力結果が白色化されないこと

最初の 2 点は 3.3 節で述べたブラインドシステム同定で問題となる点であり , 最後の 1 点が 4.4.1 項で述べたブラインド残響除去で生じる問題である .

上記 3 つの課題をすべて解決し , 現状に存在するブラインド残響除去の中で上記 3 つを解決し , 完全な残響除去が可能な方法に線形予測を用いた方法がある .

4.4.1 項で述べたも述べたとおり , 音源波形が互いに独立で同一の分布に従う場合 (independent and identically distributed = i.i.d.) は , 出力信号のみから室内インパルスの逆フィルタをブラインドに推定することができる [105–107] .

しかし , これらの方法は , 音声などの i.i.d. ではない出力信号に適用すると , 音声の自己回帰 (Auto Regressive = AR) 過程も逆フィルタリングしてしまうため , 推定された音源波形が白色化されてしまう問題がある .

この問題を解決するために，線形予測に基づく方法を用いて AR 過程を別途推定し [108,109]，白色化された音源波形を元に戻す方法である LIME アルゴリズム (Linear-predictive Multi-input Equalization) が提案されている [70,71,110]．この方法も Semi-blind-MINT と同様，音源と一番近接なマイクロホンのチャンネル番号情報を必要とし，そのマイクロホンでの信号をチャンネル 1 とする必要がある．

各マイクロホンでの入力は，

$$x_i(n) = h_i(n) * s(n) = \sum_{k=0}^{M-1} h_i(k)s(n-k) \quad (4.54)$$

と表現できる．ここで，音源信号  $s(n)$  は白色雑音  $e(n)$  を用いた有限の AR 過程で作られたと仮定する．AR 多項式は，

$$a(z) = 1 - \{a_1 z^{-1} + \dots + a_N z^{-N}\} \quad (4.55)$$

と表現でき，音源信号  $s(n)$  は，

$$\mathbf{s}(n) = \mathbf{C}^T \mathbf{s}(n-1) + \mathbf{e}(n) \quad (4.56)$$

$$\mathbf{C} = \begin{pmatrix} a_1 & 1 & 0 & \dots & 0 \\ a_2 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & & \ddots & 1 \\ a_N & 0 & \dots & \dots & 0 \end{pmatrix} \quad (4.57)$$

と表現できる．ここで， $\mathbf{s}(n) = [s(n), \dots, s(n-(N+1))]^T$ ， $\mathbf{e}(n) = [e(n), 0, \dots, 0]^T$  である．

入力信号  $\mathbf{x}_i(n)$  をフィルタ長  $L$  の予測フィルタ  $\mathbf{w}_i$  でフィルタリングし，室内インパルス除去および音源信号の白色化する．式 (4.54) を行列表現すると，

$$\mathbf{x}_i(n) = \mathbf{H}_i^T \mathbf{s}(n) \quad (4.58)$$

となる．ここで， $\mathbf{H}_i$  は  $(M+L-1) \times L$  の畳み込み行列であり，

$$\mathbf{H}_i = \begin{pmatrix} h_i(0) & 0 & \dots & 0 \\ h_i(1) & h_i(0) & \ddots & \vdots \\ \vdots & & \ddots & \\ h_i(M-1) & & & 0 \\ 0 & h_i(M-1) & & h_i(0) \\ \vdots & & \ddots & \vdots \\ 0 & \dots & 0 & h_i(M-1) \end{pmatrix} \quad (4.59)$$

と表現できる．

$\mathbf{h}_i = [h_i(0), \dots, h_i(M-1)]^T$  ,  $\mathbf{x}_i(n) = [x_i(n), \dots, x_i(n-(L-1))]^T$  , である．さらに ,  $\mathbf{x}(n) = [\mathbf{x}_1^T(n), \dots, \mathbf{x}_P^T(n)]^T$  ,  $\mathbf{H} = [\mathbf{H}_1, \dots, \mathbf{H}_P]$  , と定義すると ,

$$\mathbf{x}(n) = \mathbf{H}^T \mathbf{s}(n) \quad (4.60)$$

となる．ここで , 室内インパルスが共通零点を持たないとする , 行列  $\mathbf{H}$  がフルランクとなるためには ,

$$PL \geq (M+L-1) \quad (4.61)$$

となり , 予測フィルタ  $\mathbf{w}_i$  のフィルタ長  $L$  は ,

$$L \geq \frac{M-1}{P-1} \quad (4.62)$$

であればよい．

また , AR 多項式の次数  $N$  は ,

$$N = M+L-1 \quad (4.63)$$

となる [70] .

線形予測理論 [85] によれば , 予測残差  $\hat{e}(n)$  は ,

$$\hat{e}(n) = x_1(n) - \mathbf{x}^T(n-1)\mathbf{w} \quad (4.64)$$

$$= \mathbf{s}^T(n)\mathbf{h}_1 - \mathbf{s}^T(n-1)\mathbf{H}\mathbf{w} \quad (4.65)$$

となる． $\mathbf{w}$  は予測フィルタ行列 ,

$$\mathbf{w} = [\mathbf{w}_1^T, \dots, \mathbf{w}_P^T]^T \quad (4.66)$$

でありタップ長は  $PL$  である． $\mathbf{w}_i = [w_i(0), \dots, w_i(L-1)]^T$  ,  $i = 1, \dots, P$  である．式 (4.65) から , 予測残差の平均二乗値を最小にする  $\mathbf{w}$  は ,

$$\mathbf{w} = (\mathbf{H}^T \mathbf{E}\{\mathbf{s}(n-1)\mathbf{s}^T(n-1)\}\mathbf{H})^+ \mathbf{H}^T \mathbf{E}\{\mathbf{s}(n-1)\mathbf{s}^T(n)\}\mathbf{h}_1 \quad (4.67)$$

となる．ここで ,  $\mathbf{A}^+$  は Moore-Penrose 一般逆行列 [111] である． $\mathbf{h}_1$  を  $\mathbf{H}$  で置き換えた行列  $\mathbf{Q}$  は ,

$$\mathbf{Q} = (\mathbf{H}^T \mathbf{E}\{\mathbf{s}(n-1)\mathbf{s}^T(n-1)\}\mathbf{H})^+ \mathbf{H}^T \mathbf{E}\{\mathbf{s}(n-1)\mathbf{s}^T(n)\}\mathbf{H} \quad (4.68)$$

となる．式 (4.56) から ,

$$\mathbf{E}\{\mathbf{s}(n-1)\mathbf{s}^T(n)\} = \mathbf{E}\{\mathbf{s}(n-1)\mathbf{s}(n-1)\}\mathbf{C} \quad (4.69)$$

となる． $E\{s(n-1)s^T(n-1)}$  は正值のため， $X^T X$  で置き変えると，

$$\begin{aligned}
 \mathbf{Q} &= (\mathbf{H}^T \mathbf{X}^T \mathbf{X} \mathbf{H})^+ \mathbf{H}^T \mathbf{X}^T \mathbf{X} \mathbf{C} \mathbf{H} \\
 &= (\mathbf{X} \mathbf{H})^+ \mathbf{X} \mathbf{C} \mathbf{H} \\
 &= \mathbf{H}^T (\mathbf{H} \mathbf{H}^T)^{-1} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{X} \mathbf{C} \mathbf{H} \\
 &= \mathbf{H}^T (\mathbf{H} \mathbf{H}^T)^{-1} \mathbf{C} \mathbf{H}
 \end{aligned} \tag{4.70}$$

となる．よって，

$$\mathbf{w} = \mathbf{H}^T (\mathbf{H} \mathbf{H}^T)^{-1} \mathbf{C} \mathbf{h}_1 \tag{4.71}$$

となる．式 (4.71) より，AR 過程の次数  $N = M + L - 1$  は予測フィルタ行列  $\mathbf{w}$  に収まっていることがわかる．また，

$$\begin{aligned}
 \hat{e}(n) &= \mathbf{s}^T(n) \mathbf{h}_1 - \mathbf{s}^T(n-1) \mathbf{H} \mathbf{w} \\
 &= \mathbf{s}^T(n) \mathbf{h}_1 - \mathbf{s}^T(n-1) \mathbf{H} \mathbf{H}^T (\mathbf{H} \mathbf{H}^T)^{-1} \mathbf{C} \mathbf{h}_1 \\
 &= (\mathbf{s}^T(n) - \mathbf{s}^T(n-1) \mathbf{C}) \mathbf{h}_1 \\
 &= \mathbf{e}^T(n) \mathbf{h}_1 \\
 &= h_1(0) e(n)
 \end{aligned} \tag{4.72}$$

より，予測フィルタ  $\mathbf{w}$  によって予測残差信号  $\hat{e}(n)$  には室内残響成分  $\mathbf{H}$  は完全に取り除かれ， $e(n)$  となり白色化されていることが確認できる．

この予測誤差  $\hat{e}(n)$  から音源信号  $s(n)$  を推定するためには，式 (4.55) の AR 多項式  $a(z)$  を推定する必要がある．推定した AR 多項式の  $1/\hat{a}(z)$  逆特性によって，音源信号  $s(n)$  を回復する．

式 (4.57) の行列  $\mathbf{C}$  の特性多項式は，

$$\begin{aligned}
 f_c(\mathbf{C}, \lambda) &= -\lambda^N + a_1 \lambda^{N-1} + \cdots + a_N \\
 &= -\lambda^N \{1 - (a_1 \lambda^{-1} + \cdots + a_N \lambda^{-N})\}
 \end{aligned} \tag{4.73}$$

となる．ここで， $f_c(\mathbf{A}, \lambda) = \det(\mathbf{A} - \lambda \mathbf{I})$  は行列  $\mathbf{A}$  の特性多項式 (characteristic polynomial) である．式 (4.57) と式 (4.73) より，多項式係数  $a(z)$  は行列  $\mathbf{C}$  の特性多項式の係数に等しい．

ここで，行列  $\mathbf{Q}$  の固有値は，

$$\begin{aligned}
 \lambda(\mathbf{Q}) &= \lambda(\mathbf{H}^T (\mathbf{H} \mathbf{H}^T)^{-1} \mathbf{C} \mathbf{H}) \\
 &= \lambda(\mathbf{H} \mathbf{H}^T (\mathbf{H} \mathbf{H}^T)^{-1} \mathbf{C}) \\
 &= \lambda(\mathbf{C})
 \end{aligned} \tag{4.74}$$

となる．よって．

$$f_c(\mathbf{Q}, \lambda) = f_c(\mathbf{C}, \lambda) \quad (4.75)$$

となる．つまり，行列  $\mathbf{Q}$  の特性多項式により，AR 多項式  $\hat{a}(z)$  を推定することができる．次数の大きい行列の特性多項式は，文献 [112] の計算方法によって効率よく計算できる．推定した AR 多項式  $1/\hat{a}(z)$  を予測誤差  $\hat{e}(n)$  にフィルタリングすることにより，音源信号  $\hat{s}(n)$  を回復することができる．

実際には， $\mathbf{H}$  および  $s(n)$  はわからないため，式 (4.60)，式 (4.68) により，行列  $\mathbf{Q}$  は，

$$\mathbf{Q} = \left( \mathbf{E}\{\mathbf{x}(n-1)\mathbf{x}^T(n-1)\} \right)^+ \mathbf{E}\{\mathbf{x}(n-1)\mathbf{x}^T(n)\} \quad (4.76)$$

として計算される．

実際のアルゴリズムを以下に示す (図 4.4)．

- i) 入力信号  $\mathbf{x}$  から式 (4.77) により行列  $\mathbf{Q}$  を計算する
- ii) 行列  $\mathbf{Q}$  の最初の行から予測フィルタ  $\mathbf{w}$  を得る
- iii) 予測誤差  $\hat{e}(n)$  を式 (4.65) により計算する
- iv) 行列  $\mathbf{Q}$  の特性多項式から AR パラメータ  $a(z)$  を得る
- v) 音源信号  $s$  を予測誤差  $\hat{e}(n)$  と  $1/\hat{a}(z)$  のフィルタリングにより計算する

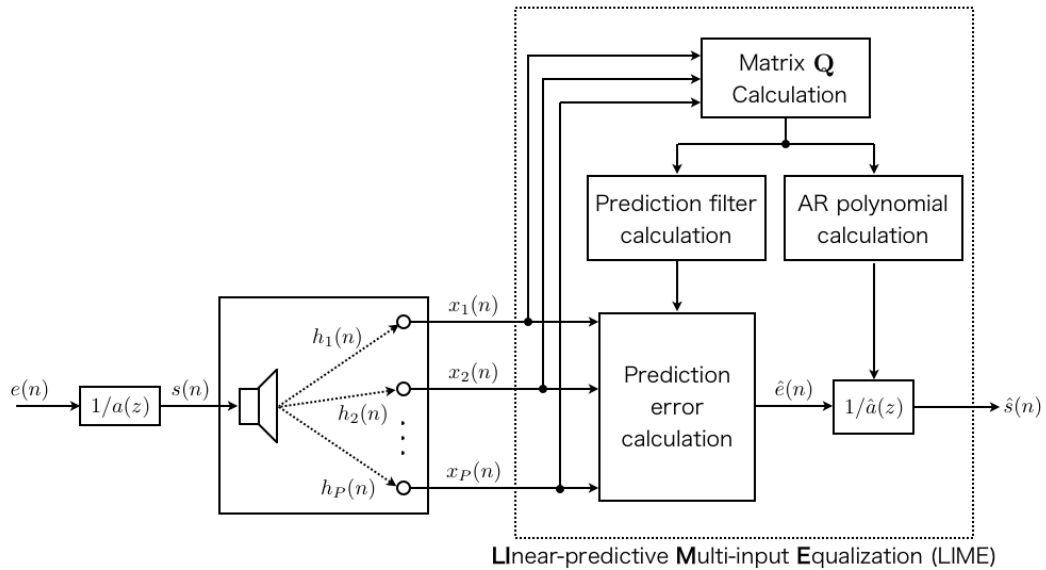


図 4.4: LIME アルゴリズムのブロック図

文献 [71] では，3840 タップの室内インパルスが畳み込まれた 8 kHz サンプリングの音声信号 2 s から 2 チャンネル～6 チャンネルを用いて残響除去を行っている．インパルス応答が長い場合，各チャンネル間に共通零点が存在することがあり，その結果，チャンネル数の少ない 2 チャンネルの場合は残響除去精度が低下する．これに対し，使用するチャンネル数を 6 チャンネルなどに増やすことにより，共通零点を回避し，精度よく残響除去ができることが報告されている．

LIME アルゴリズムの一番の特徴は，式 (4.76) の行列  $Q$  から逆フィルタとは別途に計算される特性多項式の係数がそのまま原音の AR 係数となることにある．室内インパルスの正確なタップ長  $M$  が不明な場合でも，逆フィルタ長  $L$  を式 (4.77) のように式 (4.62) よりも長め ( $= \delta$ ) に見積もっておくことにより，特性方程式から計算される AR 係数ベクトルは長く見積もったタップ分だけベクトルの末尾の値が 0 として計算されるため，正確な AR 係数が求まり，原音を精度よく抽出することができる． $\delta$  が 0 以上ならば， $\delta$  を大きくしても精度が劣化しないことが報告されている [70] ．

$$L = \frac{M - 1}{P - 1} + \delta \quad (4.77)$$

ただし，LIME アルゴリズムの特徴である行列  $Q$  から計算される特性多項式は背景雑音に非常に弱いため，雑音の存在する環境においては精度よく AR 係数を推定することができない [70, 71] ．雑音環境における音源の AR 係数推定法はブラインド残響除去の大きな課題の一つである．

## 4.5 多チャンネルブラインドシステム同定とブラインド残響除去の特徴

4.3 節，4.4 節において室内残響が畳み込まれた観測信号から原音情報を精度よく抽出するブラインドシステム同定およびブラインド残響除去について概要した．ブラインドシステム同定は系のインパルス応答を正確に求めた後 MINT により逆フィルタを形成するアプローチであり，ブラインド残響除去は多チャンネル MINT 型逆フィルタの理論に基づき，入力信号から直接逆フィルタを形成するアプローチである．

ここで，インパルス応答長を  $M$ ，観測信号のチャンネル数を  $P$  とする．

ブラインドシステム同定では，インパルス応答長を  $M$  を正確な値として推定しなければ推定精度が著しく劣化するという問題がある．文献 [113] では，正確な次数  $M$  より長く見積もって推定された応答を LIME アルゴリズムと同様，特性多項式を用いて正しい次数の推定結果に修正する方法が提案されている．しかし，この方法も背景雑音には非常に弱いという問題点もある．

ブラインドシステム同定とは、 $P$  チャンネルの観測信号から  $P$  チャンネル分のインパルス応答の係数を  $M$  個求める問題、つまり  $MP$  個の係数を推定する問題に帰着する。ブラインドシステム同定では、 $MP$  個の係数を推定する問題のため、計算に用いる行列の大きさは式 (4.32) のように  $MP \times MP$  となり、チャンネル数  $P$  が大きくなると計算に用いる行列も大きくなってしまう。

一方、ブラインド残響除去では、 $M$  の正確な値よりも大きな値として式 (4.77) から逆フィルタ長を決定することにより MINT 型逆フィルタを形成できる。ただし、このフィルタは残響成分だけではなく、原音の平均スペクトル成分である AR 係数も同時に除去してしまうため、式 (4.63) の次数  $N$  の AR 係数を推定する必要がある。式 (4.62) においては、チャンネル数  $P$  が大きくなるにともない  $PL \approx M$  として考えてよい。よって、式 (4.61) より、多チャンネルブラインド残響除去では、 $P$  チャンネルの観測信号から  $M$  個の AR 係数を推定する問題に帰着する。つまり、ブラインド残響除去の場合はチャンネル数  $P$  を増やした場合、各チャンネルの逆フィルタ長は小さくなる。よって、計算に使用するトータルの行列はチャンネル数によらず  $PL \times PL \approx M \times M$  となる。よって、ブラインド残響除去は、ブラインドシステム同定と比較すると、正確なインパルス応答長情報を必要としない点、チャンネル数を増やしても計算に用いる行列のサイズが大きくならないというメリットがあると言える。

以上の特徴から、本研究では、室内残響下における観測信号からの高精度な原音抽出という問題に対して、推定する係数が少なく、インパルス応答の正確な次数情報を必要としないブラインド残響除去による方法について検討を行う。その中でも、特に高精度で残響の除去が可能な LIME アルゴリズムに着目する。しかし、LIME アルゴリズムには文献 [70, 71] では指摘されていない問題も存在し、高精度な残響除去にはさらなるアルゴリズムの改良が必要である。以下では、LIME アルゴリズムの問題点を指摘し、その解決方法を提案する。

## 4.6 Pre-Whitening フィルタを用いた LIME アルゴリズムの高精度化

### 4.6.1 室内インパルス応答の測定

本節では、LIME アルゴリズムの問題点を指摘し、その解決方法を提案する。本項では問題点の指摘および解決方法を説明する上で必要な実環境での観測信号収録について述べる。LIME アルゴリズムは背景雑音に弱いため、以下の検討では背景雑音の混入していない観測信号を用いる必要がある。よって、観測信号は測定した室内残響に音源信号を畳み込んだものを使用することになる。よって、本項では実環境におけ

るインパルス応答収録について述べる．

包囲型マイクロホンによりインパルス応答を収録する．スピーカ配置は図 4.5 となる．本節では，使用するインパルス応答はスピーカ正面に位置する壁面 A の 20 チャンネルを用いる (図 4.6) ．

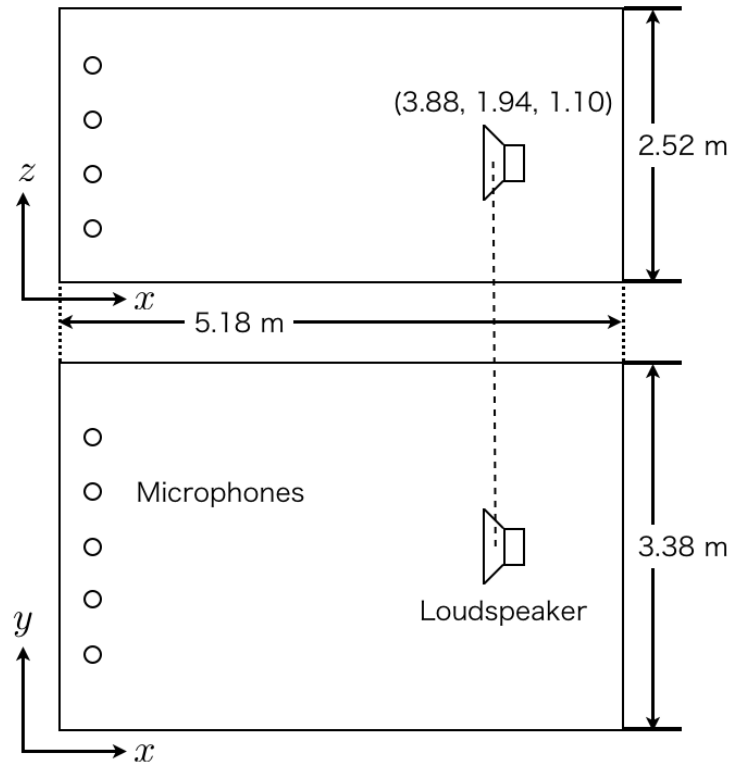


図 4.5: 室内インパルス応答測定における音源のとマイクロホンの位置関係図

スピーカはマイクロピュアの AP-5001 を使用する (図 4.7) ．音源には 48 kHz サンプリング，65536 ポイントの TSP 信号を 20 回提示する．提示音圧レベルは 92.6 dB である．20 回の同期加算後，用いた信号の逆 TSP 信号を畳み込み，インパルス応答を取得する．残響時間は 0.15 s のため，初期応答から 7200 タップをインパルス応答とする．計測したインパルス応答の例を図 4.8 に示す．これらのインパルス応答を用いて，以下の検討を進める．

#### 4.6.2 LIME アルゴリズムの問題点

雑音の混入がない SIMO モデルにおいて観測信号から高精度に残響を除去し，原音を抽出可能な LIME アルゴリズムであるが，以下のような問題がある．

LIME アルゴリズムをはじめとして，残響除去の研究の多くは，残響によって明瞭度が低下した音声信号を回復し，音声認識精度を向上させる目的で行われている．音



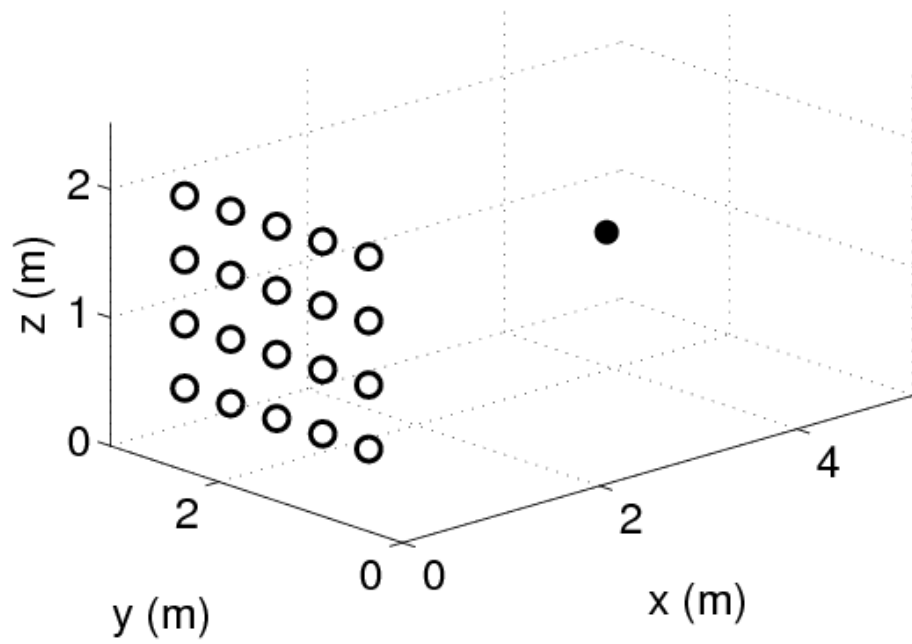


図 4.6: 室内インパルス応答測定における音源のとマイクロホン配置 (スピーカ正面 20 チャンネル) : マイクロホン, : スピーカ



図 4.7: 室内インパルス応答測定に使用したスピーカ (Micropure AP5001)

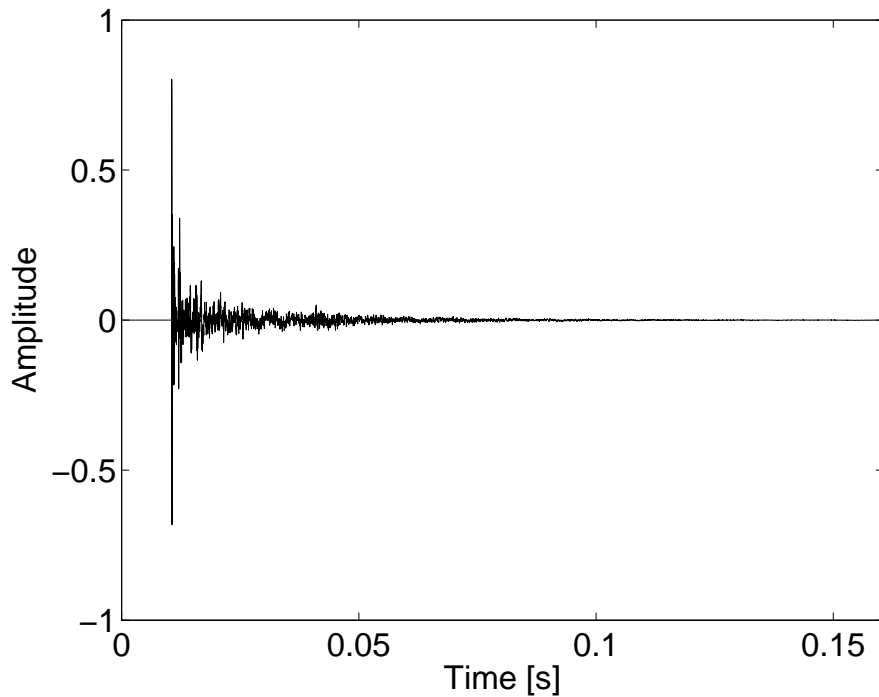


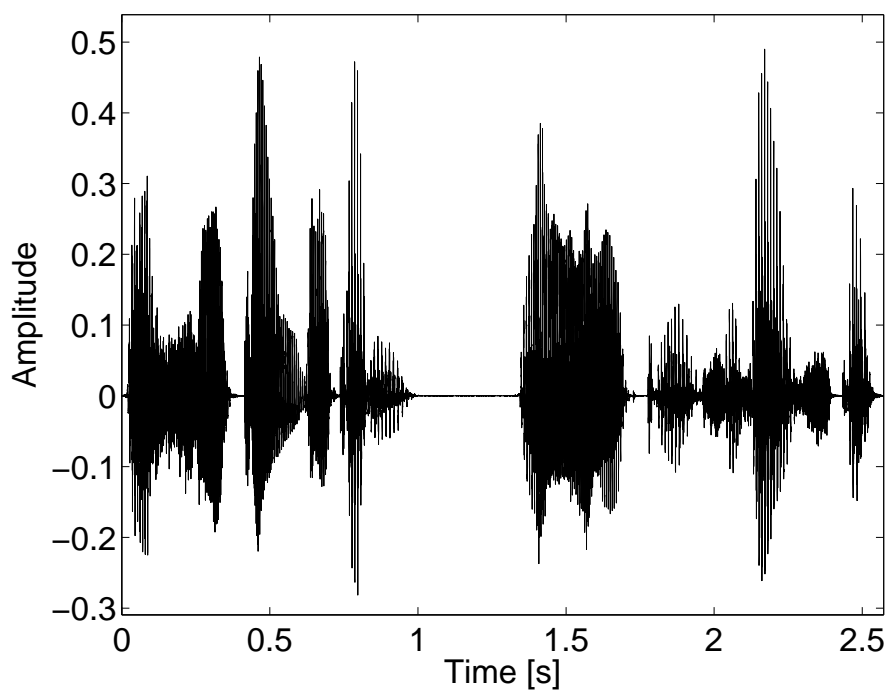
図 4.8: 室内で測定したインパルス応答の時間波形

声認識で重要となってくる帯域は 4 kHz 程度までとされており，従来研究でのアルゴリズムの検討に用いられるサンプリング周波数は 8 k ~ 16 kHz である．

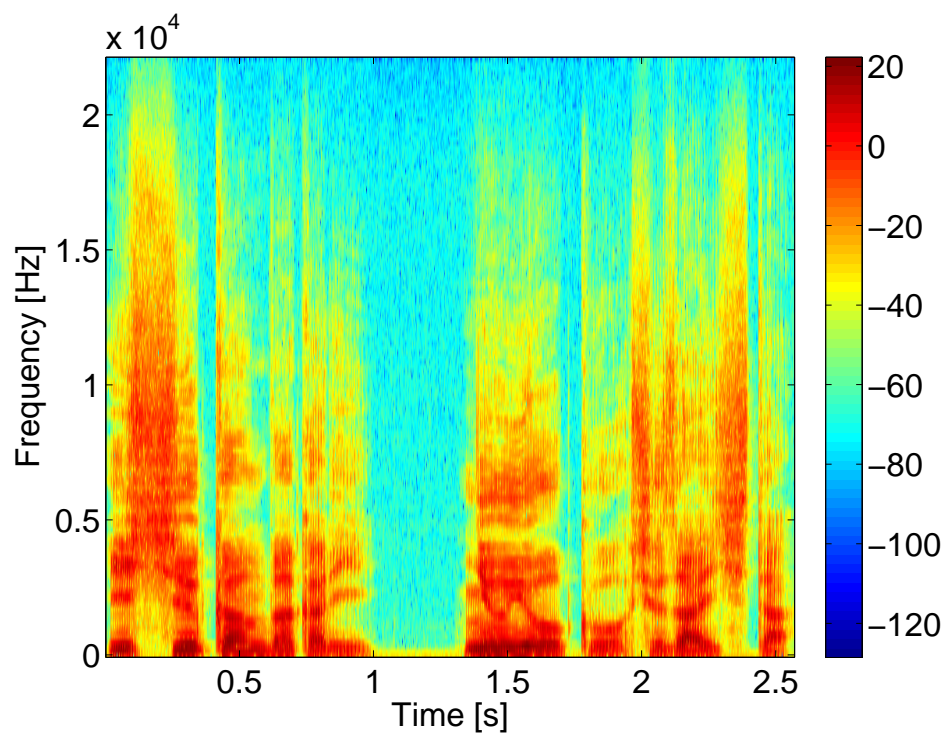
しかし，本研究ではより高精度な原音抽出を目的としており，音声認識には必要のない 16 kHz までの帯域の音声信号や，さらに高い周波数成分を含む楽器の原音情報も厳密に抽出し，高臨場感コンテンツの実現等に資することが一つの大きな目的である．図 4.9 に広帯域音声の時間係とスペクトログラムを示す．この音声信号は NTT アドバンステクノロジ株式会社の音素バランス 1000 文広帯域音声データベースの男性音声の一例である．このように，高域にも少ないながらも保存すべき情報は存在している．よって，より忠実な音場の保存，再現のためには，従来はあまり検討されてこなかった高サンプリング周波数での残響除去について検討する必要がある．

同じ残響時間の場合，サンプリング周波数が高い方がインパルス応答のタップ長は大きくなる．よって，LIME アルゴリズムでは観測信号から式 (4.76) によって計算する行列  $Q$  のサイズは大きくなる．したがって，この行列が大きくなったときの観測信号のスパース性に関する問題が生じる．以下に問題点を述べる．

実環境に存在する音声や音楽などの信号は時間軸上，周波数軸上においてスパース性をもつ信号である．スパース性とは，時間軸上や周波数上で信号のパワーがまばらになる状態を指す．LIME アルゴリズムでは行列  $Q$  を計算する際，式 (4.76) のように



(a)



(b)

図 4.9: サンプル周波数 44.1 kHz の男性音声の例  
(a) : 時間波形 (b) : スペクトログラム

時間軸的には期待値を取るため，時間軸方向のスパース性は問題とはならないが，周波数軸方向のスパース性が残響除去の精度を劣化させる問題について指摘する．

図 4.9 のサンプリング周波数 8 kHz の音声信号の平均スペクトルを図 4.10 に示す．図 4.10 のように，音声信号は低域に大きなパワーを持ち，高域に行くにしたがい，パワーが減衰している．LIME アルゴリズムでは，スパース性の中でも，高域に行くにしたがい全体的なパワーが小さくなるという性質により精度が劣化する問題がある．

式 (4.76) の行列  $Q$  の一部である観測信号の相関行列  $E\{x(n-1)x^T(n-1)\}$  に着目する．この行列は時間軸方向の期待値を取っているため，周波数的には図 4.10 のような平均スペクトルを持つことになる．残響時間が長くなればなるほど，この相関行列のサイズを大きくする必要があるが，平均スペクトルの差は細かく現れることになるため，相関行列  $E\{x(n-1)x^T(n-1)\}$  自体にも周波数軸方向にかなりエネルギーの偏った行列となる．その結果，行列の固有値の分布が大きくなり，相関行列  $E\{x(n-1)x^T(n-1)\}$  の条件数が大きくなる．条件数とは，最大固有値の絶対値と最小固有値の比率であり，値が 1 に近いほど逆行列の解の精度が高いという指標として用いることができる．したがって，条件数の増加に伴い，相関行列  $E\{x(n-1)x^T(n-1)\}$  のランクが低下する．その結果，相関行列  $E\{x(n-1)x^T(n-1)\}$  のランク数が式 (4.63) で定義される原音  $s(n)$  の AR 多項式の次数  $N$  よりも小さくなる場合は原音  $s(n)$  の AR 係数  $a(z)$  を正しく推定することができなくなる．

さらに，サンプリング周波数を高くした場合は，室内インパルス応答のタップ長が大きくなることの他に，音声信号などの場合は，周波数軸方向のパワー差がより鮮明となる．図 4.9 のサンプリング周波数 24 kHz の平均スペクトルを図 4.11 に示す．図 4.10 と図 4.11 を比較することにより，サンプリング周波数 24 kHz の場合の方が高域と低域のパワーの差が大きいことがわかる．

つまり，LIME アルゴリズムでの残響除去精度は原音の周波数軸方向のパワー差に大きく依存し，サンプリング周波数が高い場合は特にその影響が大きく現れると考えられる．逆に，白色雑音のような周波数軸方向にもパワー差が低い音源の場合は，サンプリング周波数が高い場合においても良好な残響除去が可能であると言える．具体的に説明すると，スパース性の高い信号を音声信号，スパース性の低い信号を白色雑音とし，両者を比較する．室内インパルスは図 4.8 に示した実測のインパルス応答 (サンプリング周波数 48 kHz，7200 タップ) をサンプリング周波数 24 kHz，3600 タップにダウンサンプリングしたものを使用する．

音声信号は，図 4.9 の男性音声 (サンプリング周波数 44.1 kHz) を 24 kHz にダウンサンプリングしたものを使用する．よって，マイクロホンでの観測信号はこれらの畳み込みであるため，図 4.12 のようになる．観測信号にも周波数軸上でのパワー差があることがわかる．

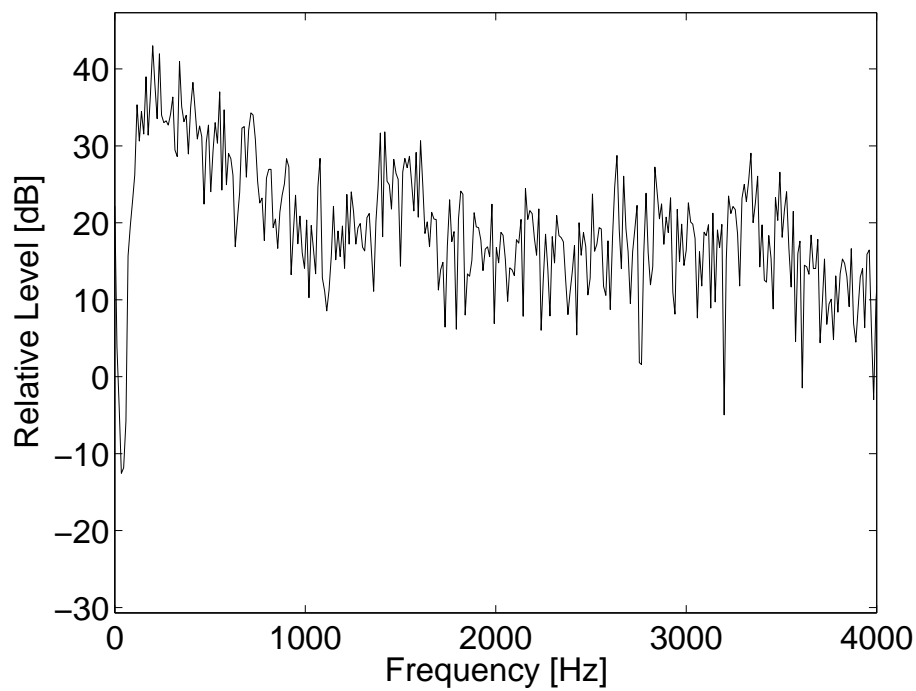


図 4.10: 男性音声 (2.5 s) の平均スペクトル (サンプリング周波数: 8 kHz)  
窓長: 1024 ポイント, オーバーラップ長: 512 ポイント

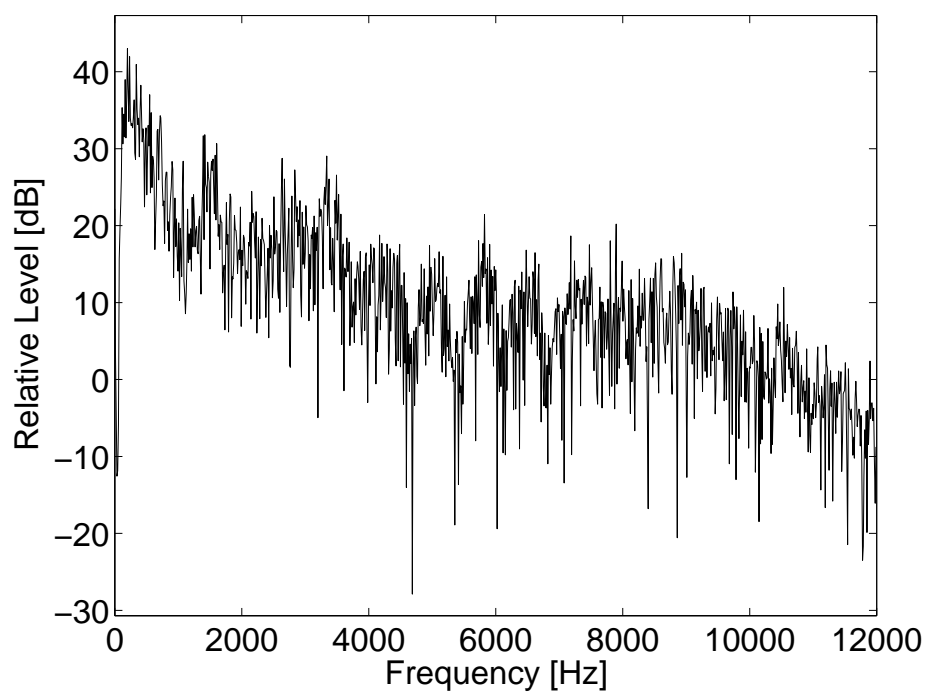
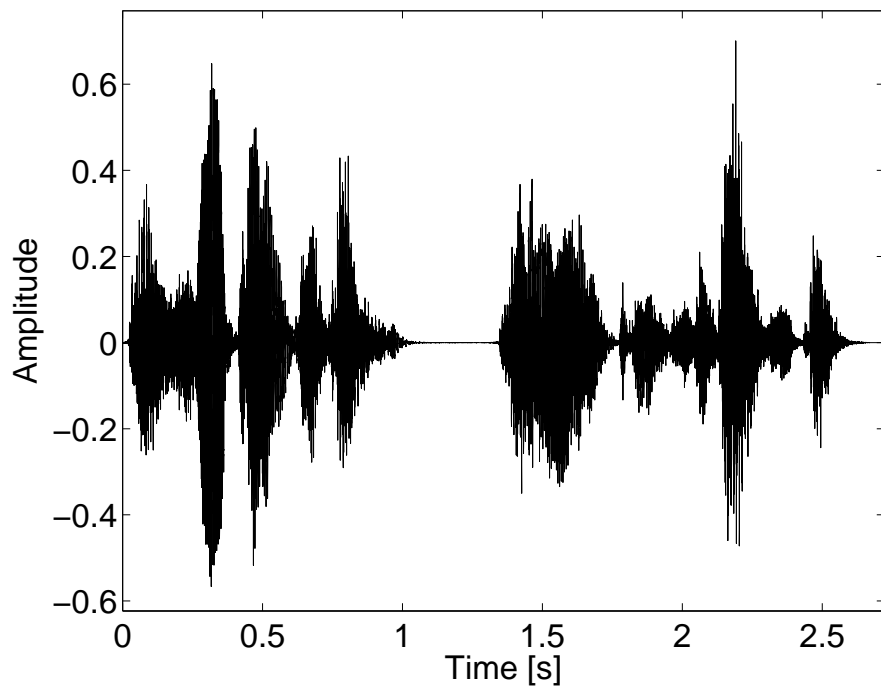
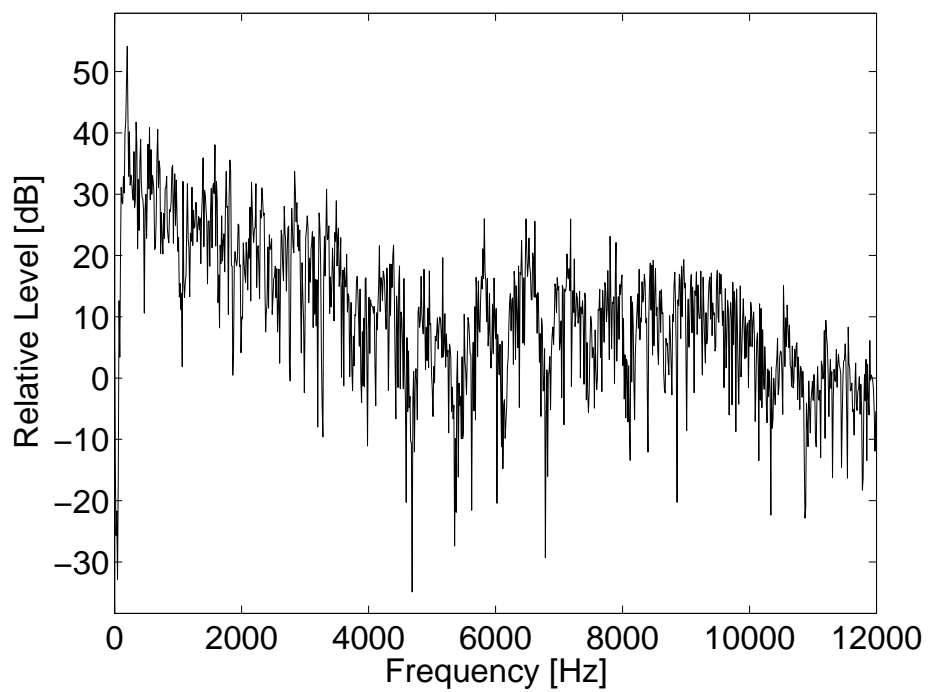


図 4.11: 男性音声 (2.5 s) の平均スペクトル (サンプリング周波数: 24 kHz)  
窓長: 1024 ポイント, オーバーラップ長: 512 ポイント



(a)



(b)

図 4.12: 原音を音声信号とした場合の室内における観測信号 (チャンネル 1)  
 (a) 波形信号 (b) 平均スペクトル (窓長: 1024 ポイント, オーバーラップ長: 512 ポイント)

白色雑音は，図 4.13(a) のように時間的に振幅を変化させたものを使用する．平均スペクトルは図 4.13(b) のように全体的にランダムであり，周波数的なスパース性は見られない．

この信号と室内インパルスを畳み込んだ観測信号は図 4.14(a)，平均スペクトルは図 4.14(b) のようになる．入力信号が音声信号のとき (図 4.12) と比べ，どの帯域も平均的にはパワーが均一であることがわかる．

これら 2 つの信号を用いて LIME アルゴリズムにおける性能を比較する．

使用するチャネル数は 20 とする．先行研究 [71] と同様，式 (4.47) の  $\delta = 10$  とすると，チャネル数  $P = 20$ ，インパルス応答長  $L = 3600$  より，逆フィルタ長  $L$  は  $L = 199$  となる．よって，この場合の原音  $s(n)$  の AR 多項式の次数  $N$  は  $N = 3600 + 199 - 1 = 3798$  タップとなる．

それぞれの相関行列  $E\{\mathbf{x}(n-1)\mathbf{x}^T(n-1)\}$  の条件数とランクを表 4.1 に示す．これらを比較すると，入力信号が音声信号の場合，ランクが 3763 と AR 多項式の次数  $N = 3798$  以下であるのに対して，入力信号が白色雑音の場合は相関行列  $E\{\mathbf{x}(n-1)\mathbf{x}^T(n-1)\}$  のランクは  $N$  と等しい値であることがわかる．

表 4.1: 相関行列  $E\{\mathbf{x}(n-1)\mathbf{x}^T(n-1)\}$  の条件数とランク

	condition number	rank
LIME (Speech)	$1.0 \times 10^{20}$	3763
LIME (White noise)	$9.3 \times 10^{18}$	3798

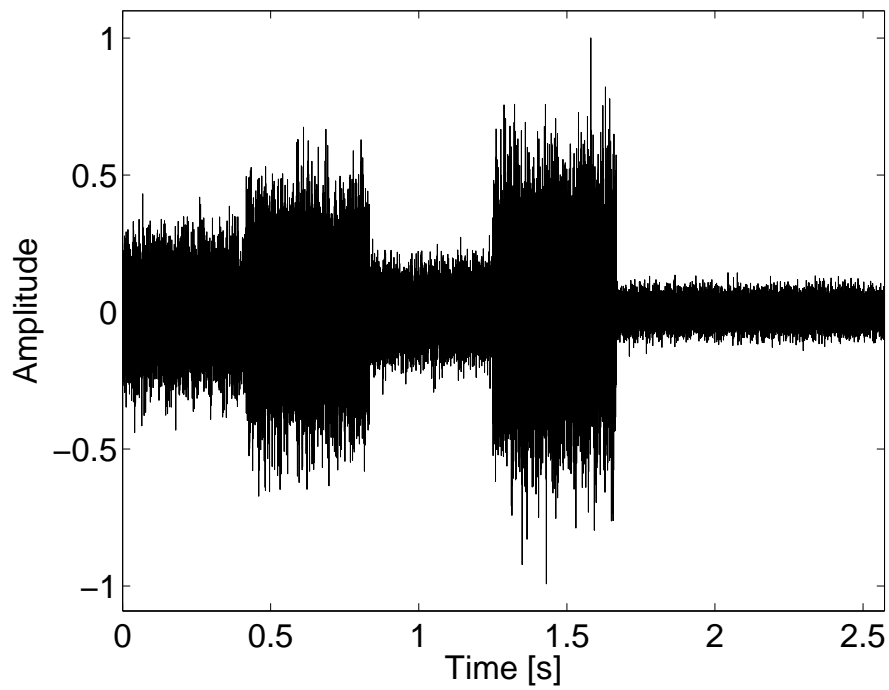
したがって，この相関行列  $E\{\mathbf{x}(n-1)\mathbf{x}^T(n-1)\}$  を用いて LIME アルゴリズムを適用すると，入力信号が音声信号の場合は行列  $\mathbf{Q}$  の特性多項式の正確な計算が不能となる．一方，入力信号が白色雑音の場合は図 4.15 のように精度よく残響除去された信号を得ることができる．

このように，入力信号が音声信号のようなスパース性のある信号の場合，LIME アルゴリズムは残響が長い場合や高サンプリング周波数の場合では正確に動作しない．

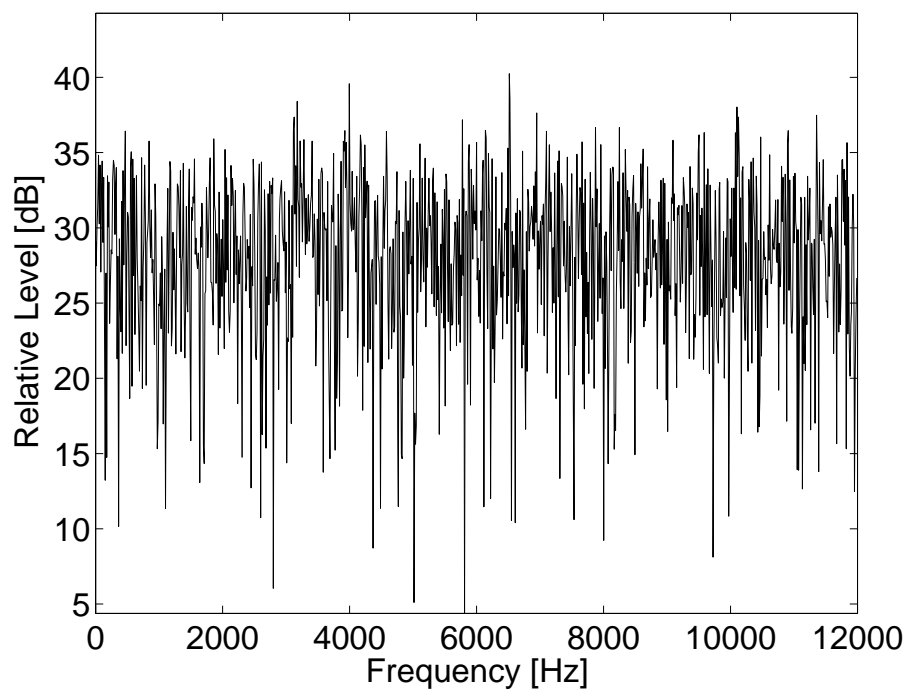
### 4.6.3 White-LIME の提案

4.6.2 項での検討を踏まえ，本項において上記の問題の解決法を提案する．

入力信号にスパース性があるならば，LIME アルゴリズムを適用する前にあらかじめそのスパース性を緩和することができればよいと考える．スパース性を緩和し，入力信号が白色化されていれば，LIME アルゴリズムが正確に動作することは 4.6.2 項で確認した通りである．



(a)

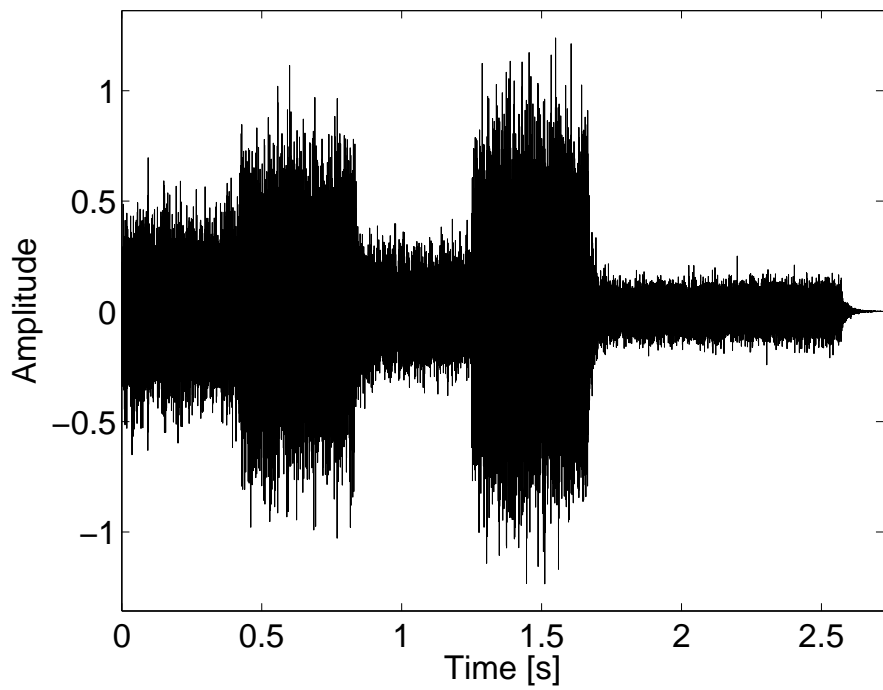


(b)

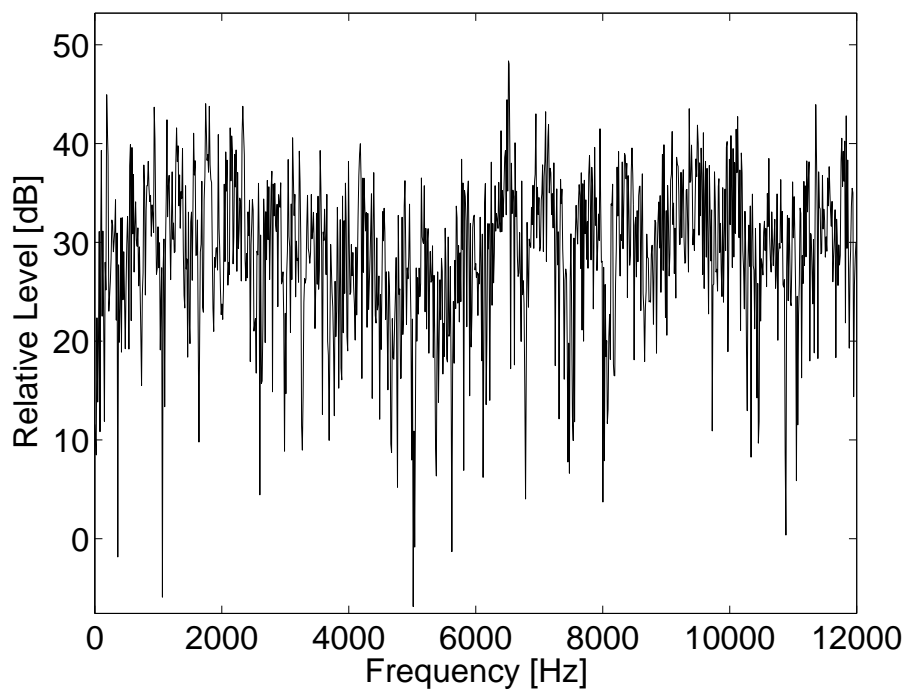
図 4.13: 原音に用いる白色雑音

(a) 波形信号 (b) 平均スペクトル (窓長: 1024 ポイント, オーバーラップ長: 512 ポイント)





(a)



(b)

図 4.14: 原音を白色雑音とした場合の室内における観測信号 (チャンネル 1)  
 (a) 波形信号 (b) 平均スペクトル (窓長: 1024 ポイント, オーバーラップ長: 512 ポイント)

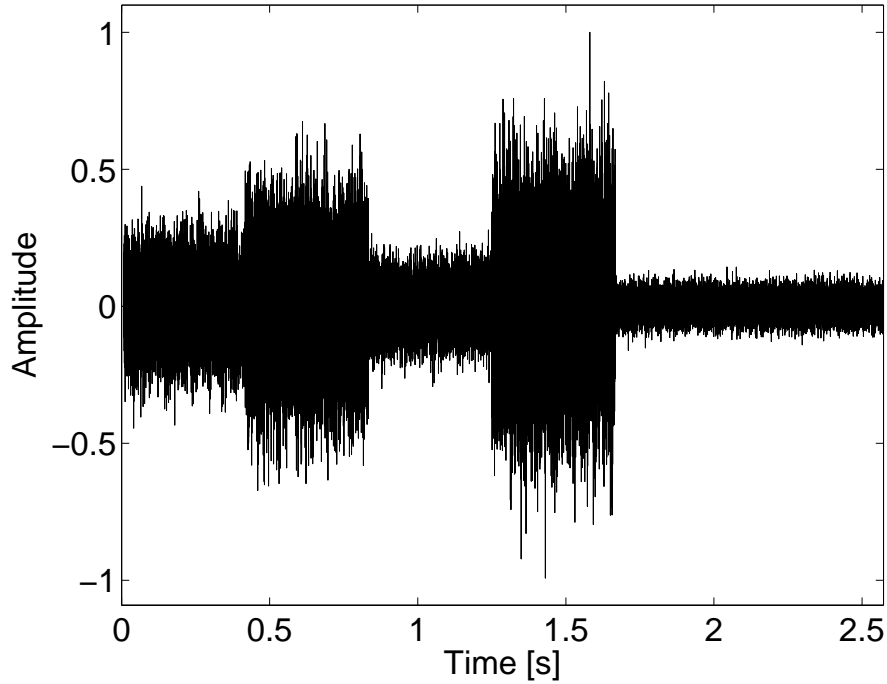


図 4.15: LIME アルゴリズムによる出力信号 (白色雑音を入力信号とした場合)

入力信号を白色化する方法としては，LIME アルゴリズムでも用いられている線形予測に着目する．LIME アルゴリズムでは，音源信号を式 (4.56) のように白色雑音と予測係数 (AR 係数) から記述している．これは，信号  $x(n)$  は予測係数  $a(n)$  と白色雑音  $e(n)$  から以下のように記述できることに基づく．

$$x(n) = \sum_{k=1}^N a(k)x(n-k) + e(n) \quad (4.78)$$

この式を書き換えると，

$$e(n) = x(n) - \sum_{k=1}^N a(k)x(n-k) \quad (4.79)$$

となる．よって式 (4.79) から，信号  $x(n)$  の予測係数  $a(n)$  を信号  $x(n)$  に畳み込むことによって信号  $x(n)$  から白色雑音  $e(n)$  を生成することができる．

線形予測のこの考え方に基づき，LIME アルゴリズムを高精度化する方法を提案する．式 (4.79) を観測信号  $x_i(n)$  に適用することにより，観測信号を白色化することができる．白色化された観測信号  $x'_i(n)$  を用いて LIME アルゴリズムを適用することにより，入力信号のスパース性を緩和できる．しかし，各チャンネルの AR 係数をそれぞれ求め，それぞれ白色化した場合，すべてのチャンネルに別々のフィルタが畳み込まれるため，LIME アルゴリズムで除去するフィルタ長が長くなる．つまり畳み込んだ AR

係数の次数分だけ室内インパルスが長くなる．それに加え，それらの AR 係数からフィルタが共通零点を有する場合，LIME アルゴリズムではそのフィルタの効果を取り除くことができなくなる．

そこで，全チャンネルに共通した白色化フィルタを用いる方法を提案する．LIME アルゴリズムでは，室内残響と音源信号の予測係数を同時に取り除いた後，予測係数のみ別途推定し，元に戻すことができる．つまり，全チャンネル共通に畳み込んだフィルタの効果は行列  $Q$  の特性多項式によって別途推定され，取り除くことが可能である．

以下，提案する具体的な方法を述べる．

まず最初に，全チャンネルに共通の白色化フィルタを観測信号  $x_i(n)$  から計算する．白色化フィルタには文献 [114, 116] で用いられている各入力信号の AR 係数の期待値を用いる．文献 [114] では各信号を別々に白色化した信号とこの係数を用いて残響成分を緩和できると報告している．各観測信号から式 (4.80) の予測係数  $b_i(n)$  は Yule-Walker 方程式を Levinson-Durbin のアルゴリズム [115] を用いて解くことによって算出する．

$$x_i(n) = \sum_{k=1}^N b_i(k)x_i(n-k) + x'_i(n) \quad (4.80)$$

$$E\{b_i(z)\} = 1 - \{E\{b_i(1)\}z^{-1} + \cdots + E\{b_i(N)\}z^{-N}\} \quad (4.81)$$

このフィルタ  $E\{b_i(z)\}$  を各観測信号  $x_i(n)$  に畳み込むことにより，各観測信号は完全ではないがほぼ白色化される．この信号を用いて LIME アルゴリズムを適用する．LIME アルゴリズムより得られた信号は白色化フィルタ  $E\{b_i(z)\}$  が畳み込まれたものが出力されるため，最後に  $E\{b_i(z)\}$  で逆畳み込みすることによって，原音信号を回復することができる．つまりこの方法は，LIME アルゴリズムで問題となる信号のスパース性を LIME アルゴリズム内において問題とならないよう観測信号をあらかじめ白色化し，LIME アルゴリズム後に白色化を元に戻すことが提案の要である．

提案法のブロック図を図 4.16 に示す．この提案法を用いることによって，スパース性のある信号を用いた場合においても高精度に残響を取り除くことができる．提案方法を White-LIME と呼ぶ．

サンプリング周波数 24 kHz，室内インパルス 3600 タップ，20 チャンネルを用いた場合を例に信号処理の流れを説明する．観測信号  $x_i(n)$  は図 4.9(a)，平均スペクトルは図 4.10(b) となる．観測信号から計算した白色化フィルタ  $E\{b_i(z)\} = b(n)$  は図 4.17(a)，平均スペクトルは図 4.17(b) となる．白色化フィルタは各観測信号  $x_i(n)$  を白色化するため，平均スペクトルは高域から低域にパワーが減衰しており，観測信号の傾向と逆であることがわかる．白色化された後の観測信号  $b(n) * x_i(n)$  は図 4.18(b)，平均スペクトルは図 4.18(b) となり，白色化信号により観測信号が白色化されていることが確認できる．

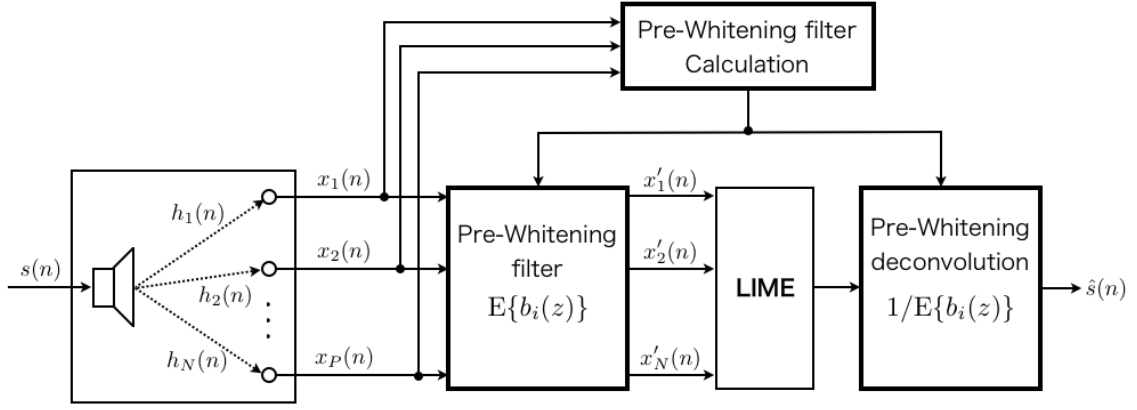


図 4.16: 提案する White-LIME のブロック図

白色化を導入しない場合 (LIME) とした場合 (White-LIME) の相関行列  $E\{\mathbf{x}(n-1)\mathbf{x}^T(n-1)\}$  の条件数とランクを表 4.2 に示す．白色化により，相関行列  $E\{\mathbf{x}(n-1)\mathbf{x}^T(n-1)\}$  のランクが AR 多項式の次数  $N = 3798$  と等しいことが確認できる．つまり，白色化により原音  $s(n)$  の AR 係数  $a(z)$  を精度よく計算できることを示す．

表 4.2: 相関行列  $E\{\mathbf{x}(n-1)\mathbf{x}^T(n-1)\}$  の条件数とランク

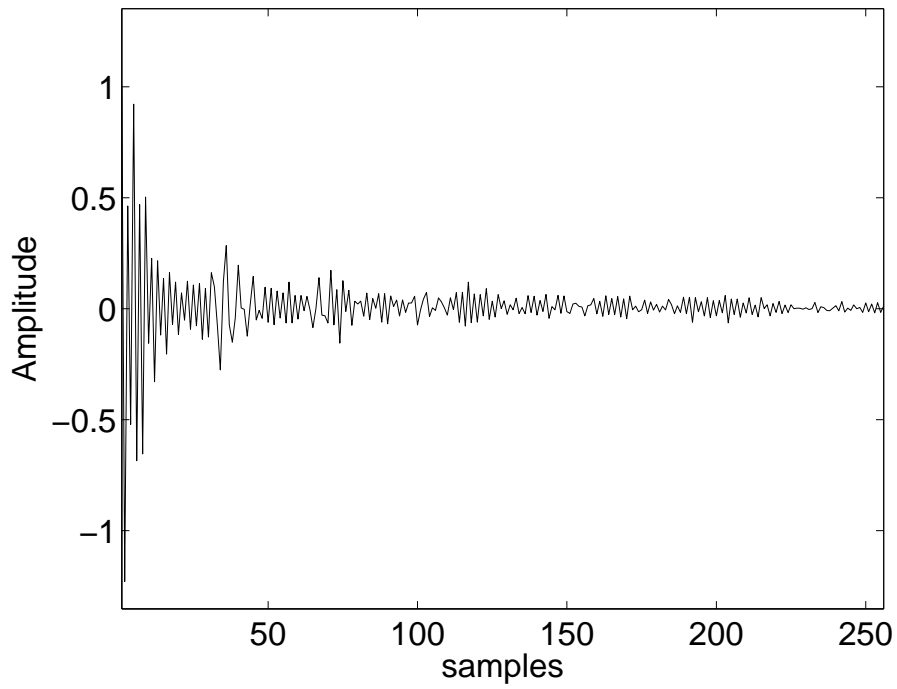
	condition number	rank
LIME (Speech)	$1.0 \times 10^{20}$	3763
White-LIME (Speech)	$6.2 \times 10^{18}$	3798

この信号を用いて LIME アルゴリズムを適用することにより得られる信号  $b(n) * y(n)$  は図 4.19(a)，平均スペクトルは図 4.19(b) となる．この信号を白色化フィルタで逆畳み込みした信号  $y(n)$  を図 4.20 に示す．逆畳み込みにより LIME アルゴリズム適用前に畳み込んだ白色化フィルタの影響を完全に除去できることが確認できる．また，原音信号，チャンネル 1 における観測信号，出力信号のスペクトルグラムを図 4.21(a)，4.21(b)，4.21(c) に示す．24 kHz の高サンプリング時においても良好に残響を除去し，原音を高精度に抽出できていることがわかる．

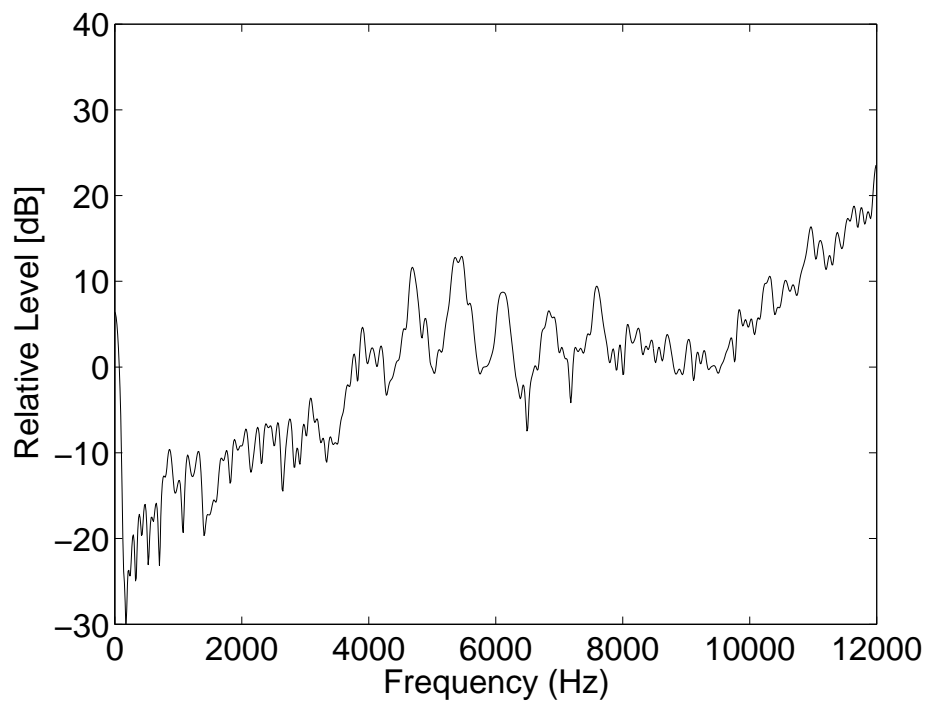
#### 4.6.4 チャネル数の違いによる性能評価

4.6.3 項において，White-LIME によって高サンプリング周波数の場合でも良好な残響除去が可能であることを示した．

ここでは，チャネル数の違いによる原音情報抽出精度について評価を行う．

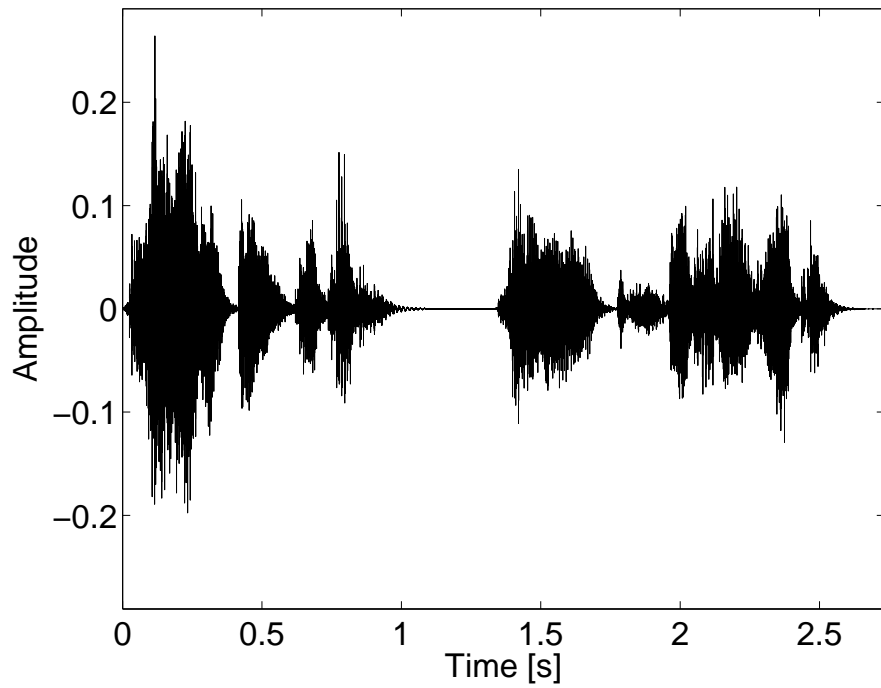


(a)

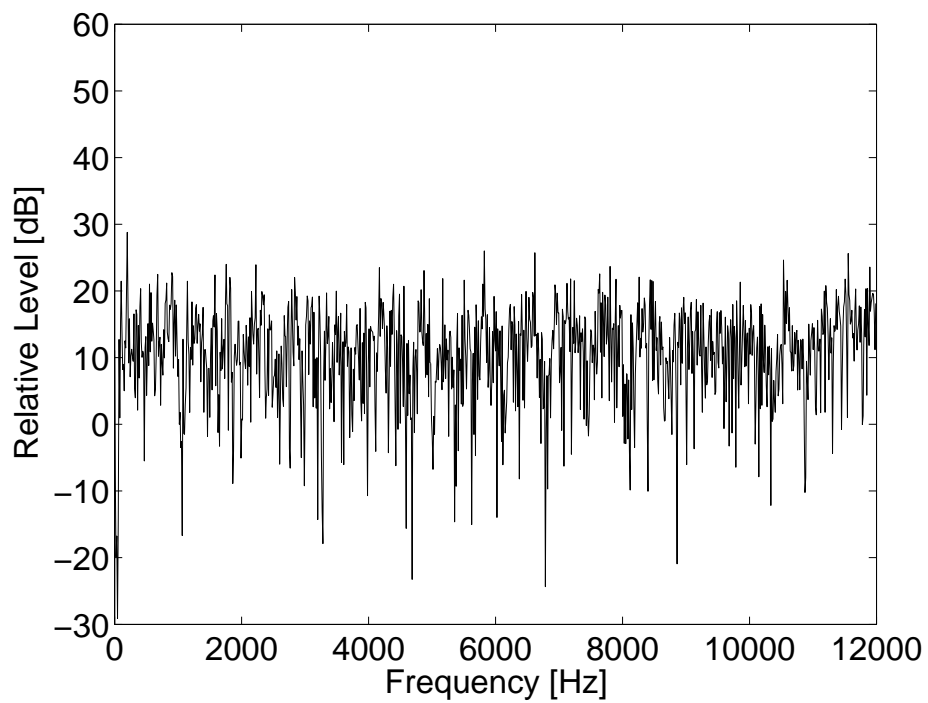


(b)

図 4.17: Pre-Whitening フィルタ  $b(n) = E\{b_i(z)\}$   
 (a) 波形信号 (b) 平均スペクトル (窓長: 1024 ポイント)



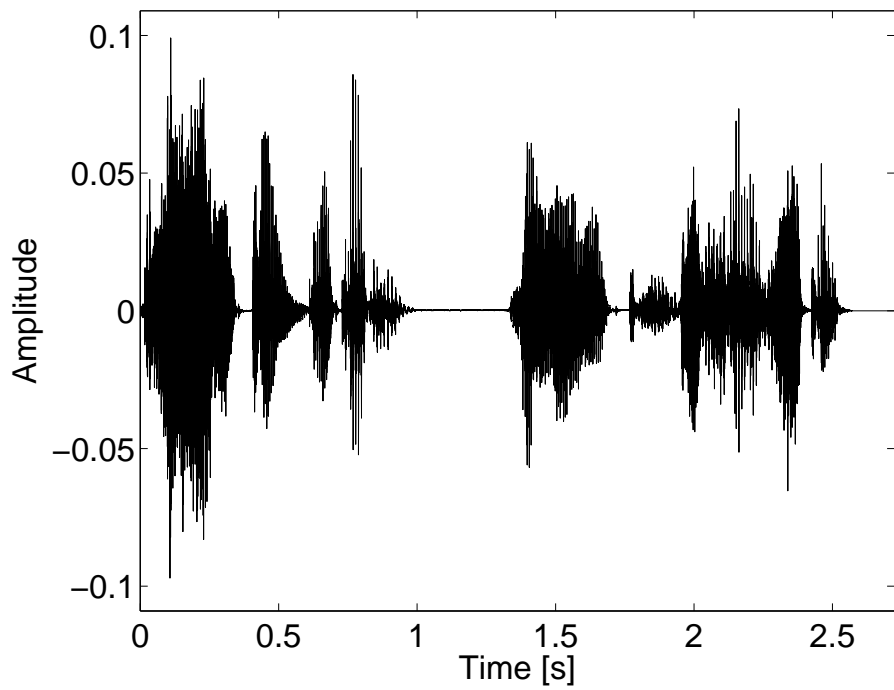
(a)



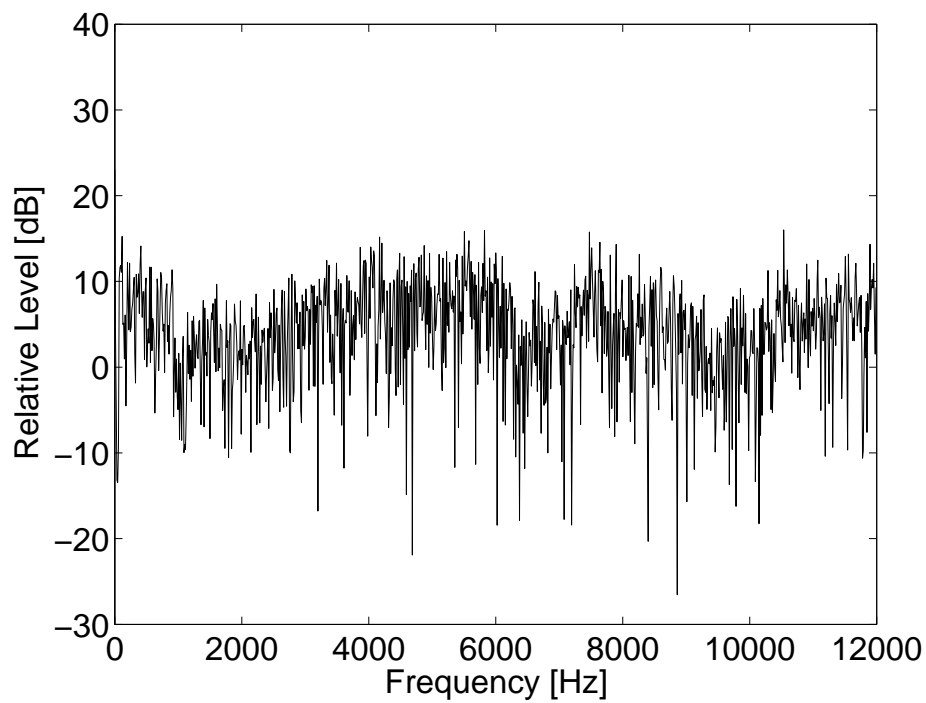
(b)

図 4.18: 白色化された観測信号  $b(n) * x_i(n)$

(a) 波形信号 (b) 平均スペクトル (窓長: 1024 ポイント, オーバーラップ長: 512 ポイント)



(a)



(b)

図 4.19: LIME 適用後の出力信号  $b(n) * y(n)$

(a) 波形信号 (b) 平均スペクトル (窓長: 1024 ポイント, オーバーラップ長: 512 ポイント)

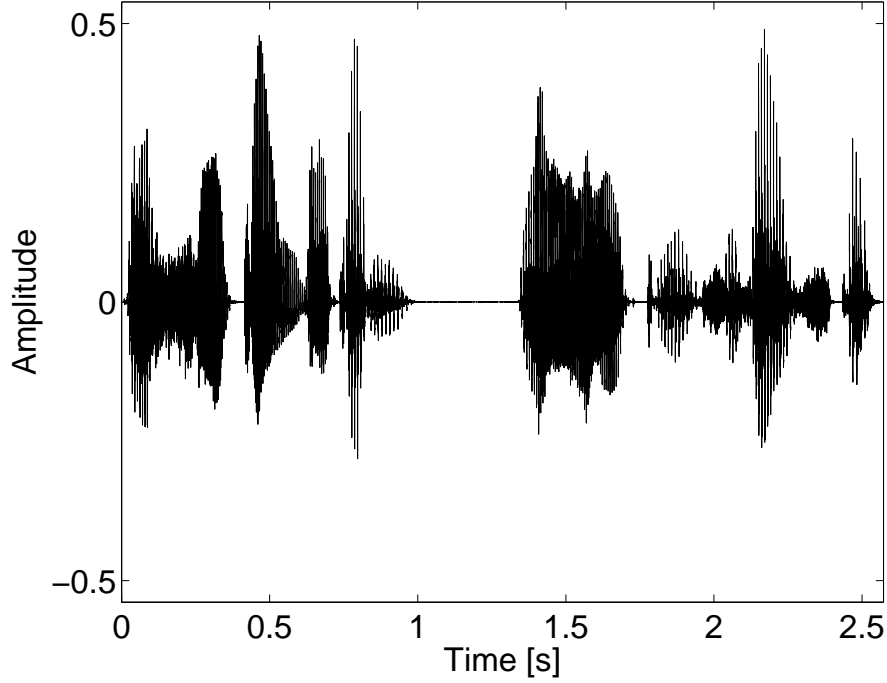


図 4.20: White-LIME による出力信号  $y(n)$

用いたチャネル数は壁面 1 面の 6, 10, 20 チャネルである．6, 10 チャネルのマイクロホン配置を図 4.22, 図 4.23 に示す．20 チャネルの配置は図 4.6 の通りである．

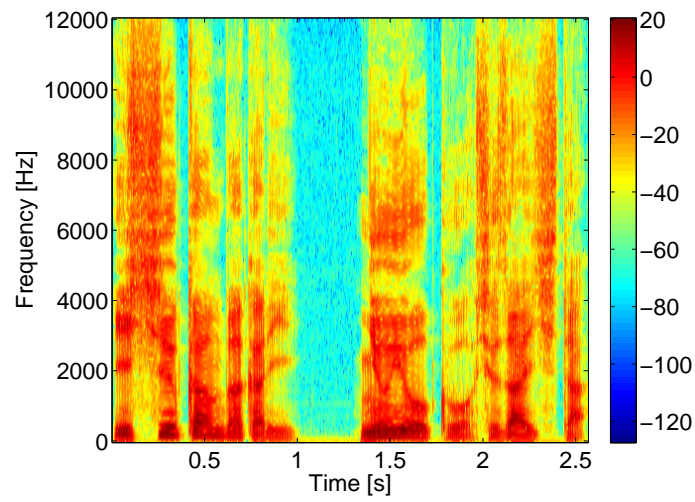
原音には図 4.9 の音声信号 (サンプリング周波数 24 kHz) を用いる．原音情報抽出の評価値には Signal to Distortion Ratio (SDR) と Spectral Distortion (SD) を用いる．SDR は値が大きいほど, SD は値が小さいほど原音との歪みが少ないことを表す指標である． $|P(f)|$  は信号の短時間フーリエ変換の値であり, 窓関数にはハミング窓 512 ポイントを用いる．SD は全時間の平均値として算出する．

$$\text{SDR} = 10 \log_{10} \left( \frac{\sum |s(n)|^2}{\sum |s(n) - \hat{s}(n)|^2} \right) \quad (4.82)$$

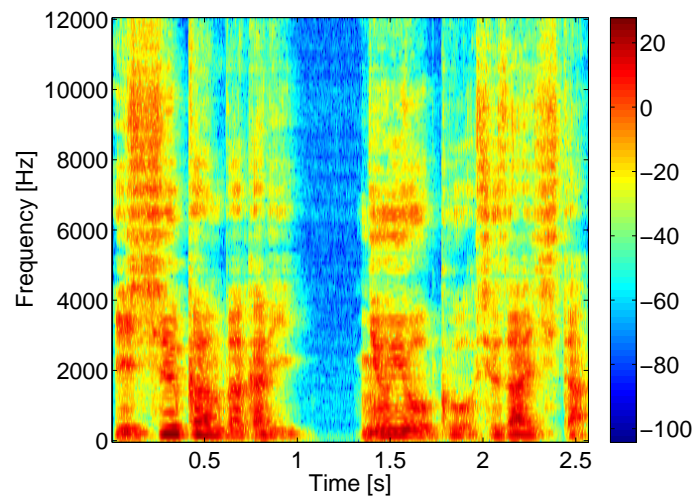
$$\text{SD} = \sqrt{\frac{1}{F} \sum_{f=0}^{F-1} \left( 20 \log |P(f)| - 20 \log |\hat{P}(f)| \right)^2} \quad (4.83)$$

表 4.3 に, 各パターンにおける AR 多項式の次数  $N$  と LIME, White-LIME それぞれの相関行列  $E\{\mathbf{x}(n-1)\mathbf{x}^T(n-1)\}$  のランク, 原音  $s(n)$  とチャネル 1 の観測信号  $x_1(n)$  との SDR, SD および原音  $s(n)$  と回復信号  $y(n)$  との SD, SDR を示す．「×」はランクの低下により回復信号を算出不可能であることを表す．また, White-LIME でのチャネル数の違いによる SDR と SD の結果のグラフを図 4.24 に示す．

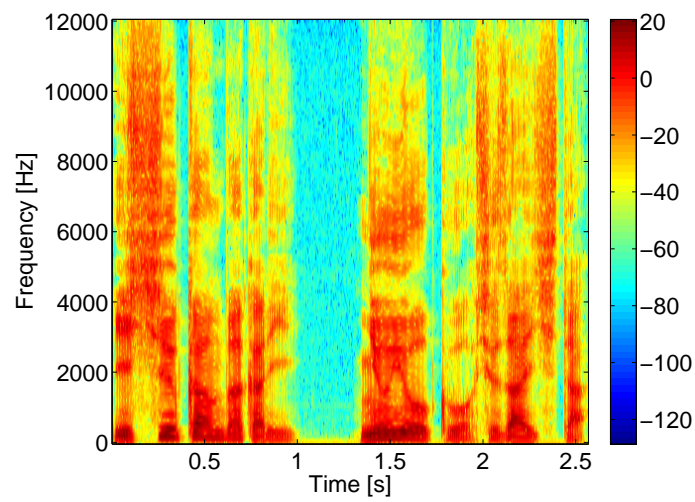




(a)



(b)



(c)

図 4.21: 原音信号, 観測信号および White-LIME による出力信号のスペクトログラム  
(a) 原音信号  $s(n)$  (b) チャンネル 1 での観測信号  $x_1(n)$  (c) 出力信号  $y(n)$

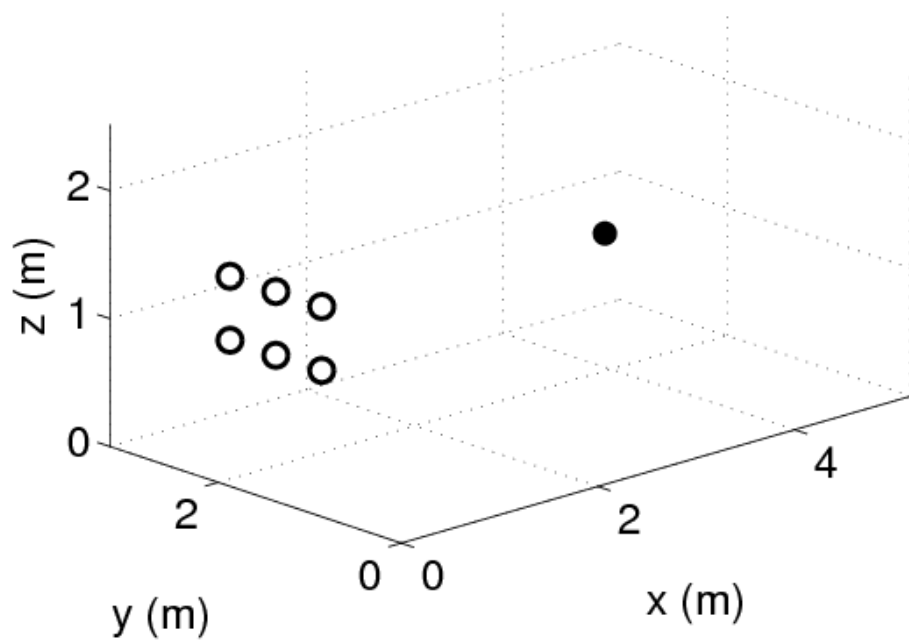


図 4.22: 室内インパルス応答測定における音源とマイクロホン配置  
(スピーカ正面 6 チャンネル)

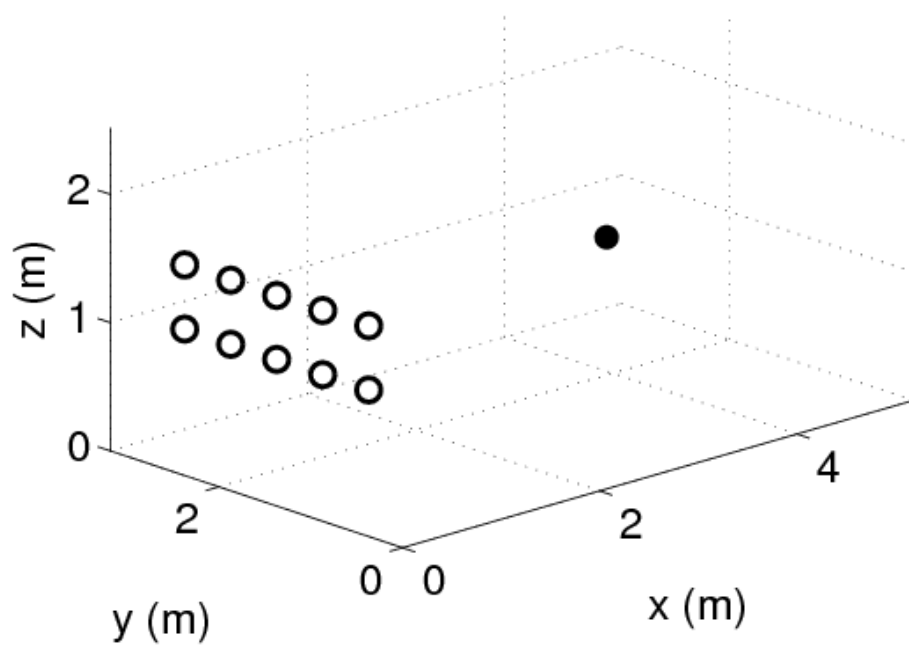


図 4.23: 室内インパルス応答測定における音源とマイクロホン配置  
(スピーカ正面 10 チャンネル)

表 4.3: チャンネル数の違いによる AR 係数  $a(z)$  の次数, LIME および White-LIME の  
相関行列  $E\{\mathbf{x}(n-1)\mathbf{x}^T(n-1)\}$  のランク, および残響除去性能評価

	order of $a(z)$	rank	SDR [dB]	SD [dB]
received signal $x_1(n)$			4.7	8.0
LIME (6 ch)	4328	4265	×	×
LIME (10 ch)	4008	3938	×	×
LIME (20 ch)	3763	3798	×	×
White-LIME (6 ch)	4328	4328	25.5	1.21
White-LIME (10 ch)	4008	4008	48.0	0.29
White-LIME (20 ch)	3798	3798	51.4	0.05

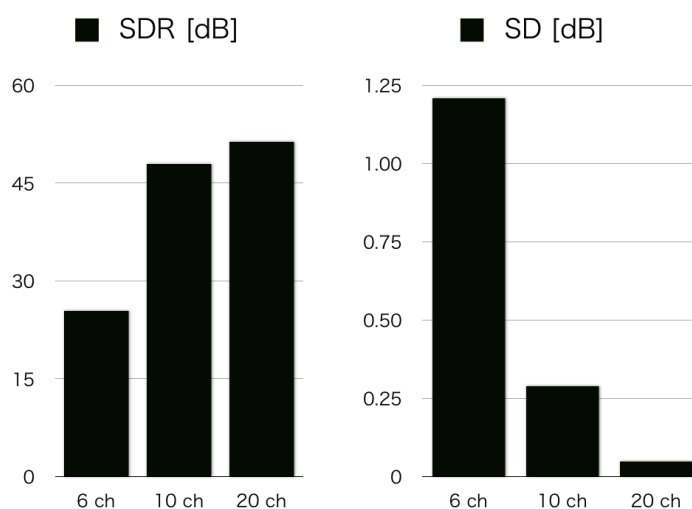


図 4.24: 原音が音声信号の場合における White-LIME のチャンネル数の違いによる性能  
評価 (サンプリング周波数: 24 kHz)

表 4.3 および図 4.24 の結果から，従来法である LIME では回復信号が算出不能であるのに対して，White-LIME では良好な残響除去が可能であることがわかる．チャンネル数を増やすことにより精度が向上している点は先行研究 [71] と同様である．

#### 4.6.5 高サンプリング周波数・高帯域信号を用いた場合の性能評価

本項では，音声信号よりもさらに高サンプリング周波数の場合，即ちサンプリング周波数 44.1 kHz の広帯域信号の場合の検討を行う．音楽信号には音楽信号は RWC 研究用音楽データベース [117] のクラシック (サンプリング周波数 44.1 kHz) 2 s を用いる．音楽信号の波形および平均スペクトルを図 4.25(a)，4.25(b) に示す．入力信号が音声信号の場合のインパルス応答は 3600 タップ，音楽信号の場合は 6615 タップとなる．6615 タップの室内インパルスが畳み込まれた音楽信号 (= 観測信号) と平均スペクトルを図 4.26(a)，4.26(b) に示す．20 チャンネル使用時の出力結果を図 4.27 に示す．サンプリング周波数 44.1 kHz においても White-LIME により高精度な残響除去が可能となる．また，SDR および SD の結果を 4.28 に示す．観測信号での SDR および SD は SDR が  $-7.5$  dB，SD が 8.9 dB であった．

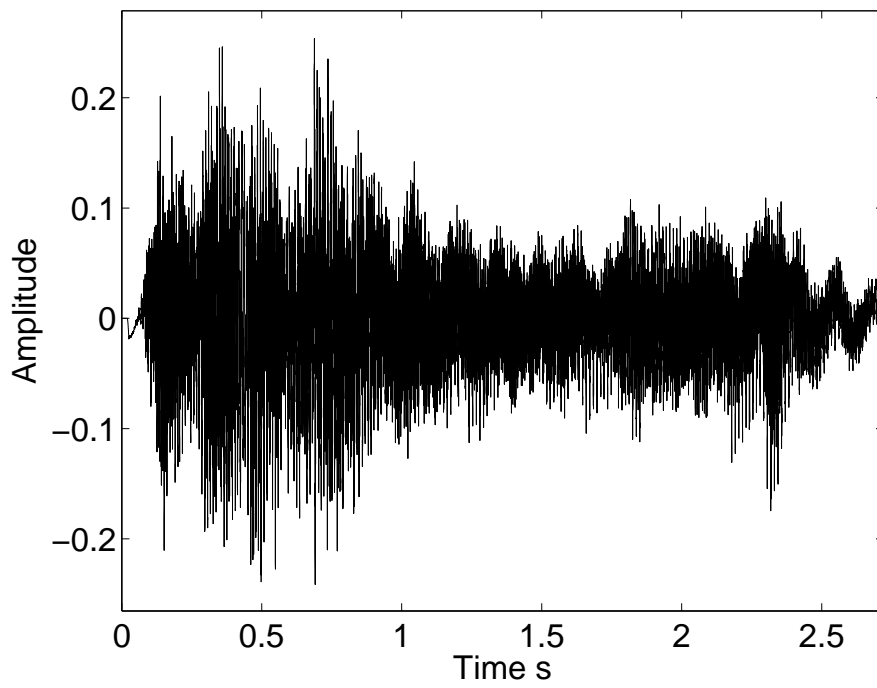
以上の結果から，サンプリング周波数 44.1 kHz の場合においても，チャンネル数を増やすことにより原音抽出精度が向上することが確認できる．

#### 4.6.6 マイクロホン配置の違いによる性能評価

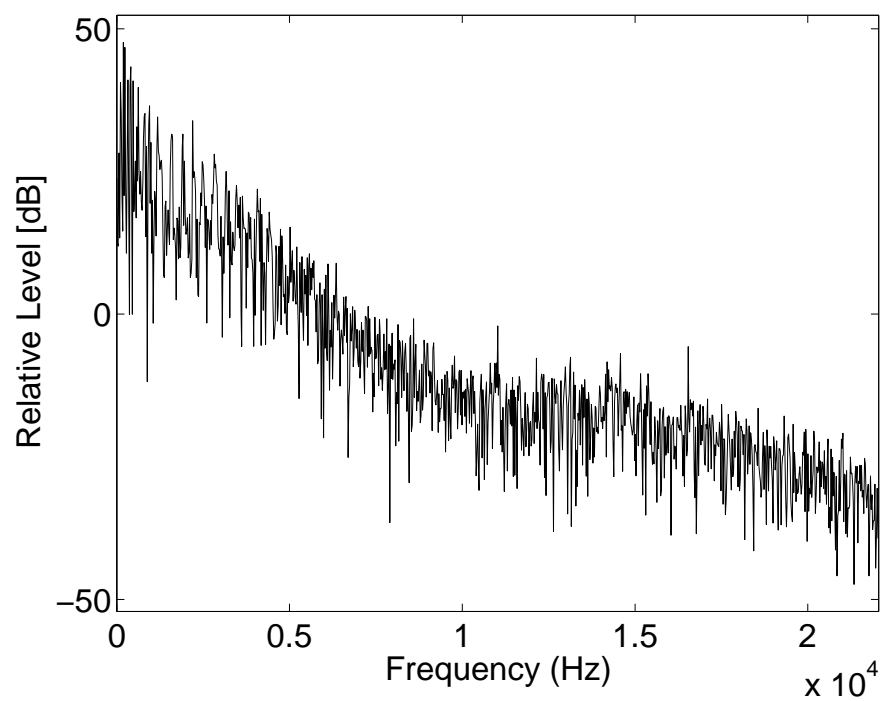
これまでの検討では，図 4.6 のように残響除去に用いるマイクロホンは 1 ウェイスピーカの正面方向の 20 ch を使用している．包囲型マイクロホンアレイは音源の各方向にマイクロホンが配置しているため，音源の正面以外からの音も收音が可能である．よって最後に本項では，マイクロホン配置の違いによる性能評価を行う．

図 4.6 のようにスピーカの正面を用いる場合．そこで，図 4.29 のようにスピーカの横側を用いる場合 (パターン 2)，後方を用いる場合 (パターン 3)，前方と横面の 2 面を用いる場合 (パターン 4) およびスピーカの前後左右の 4 面を用いる場合 (パターン 5) のとし，計 5 パターンの残響除去の性能評価を行う．シミュレーション条件はサンプリング周波数 24 kHz，原音には男性音声信号を用いた．条件は表 4.3 での検討と同様である．性能評価には SDR および SD を用いる．結果を図 4.33 に示す．

図 4.33 の結果から，20 ch を用いた場合ではどのパターンも残響除去精度は等しいことがわかる．スピーカは方向によって異なる放射指向特性を持つ．この結果は，異なる放射指向特性を持つ入力信号を用いた場合においても高精度な残響除去精度が可能であることを示している．



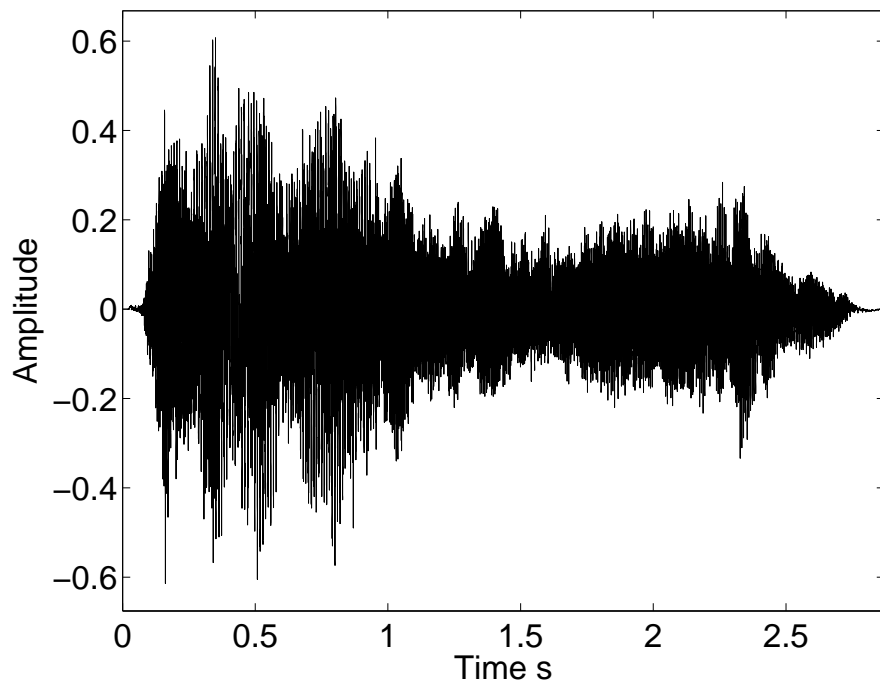
(a)



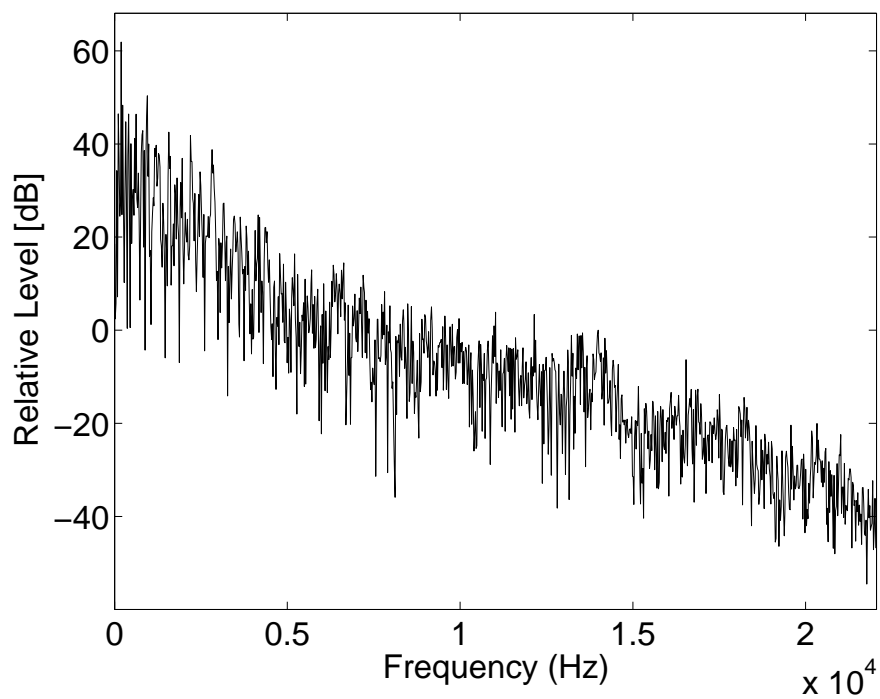
(b)

図 4.25: 原音  $s(n)$  に用いる音楽信号

(a) 波形信号 (b) 平均スペクトル (窓長: 1024 ポイント, オーバーラップ長: 512 ポイント)



(a)



(b)

図 4.26: チャンネル 1 における観測信号  $x_1(n)$

(a) 波形信号 (b) 平均スペクトル (窓長: 1024 ポイント, オーバーラップ長: 512 ポイント)

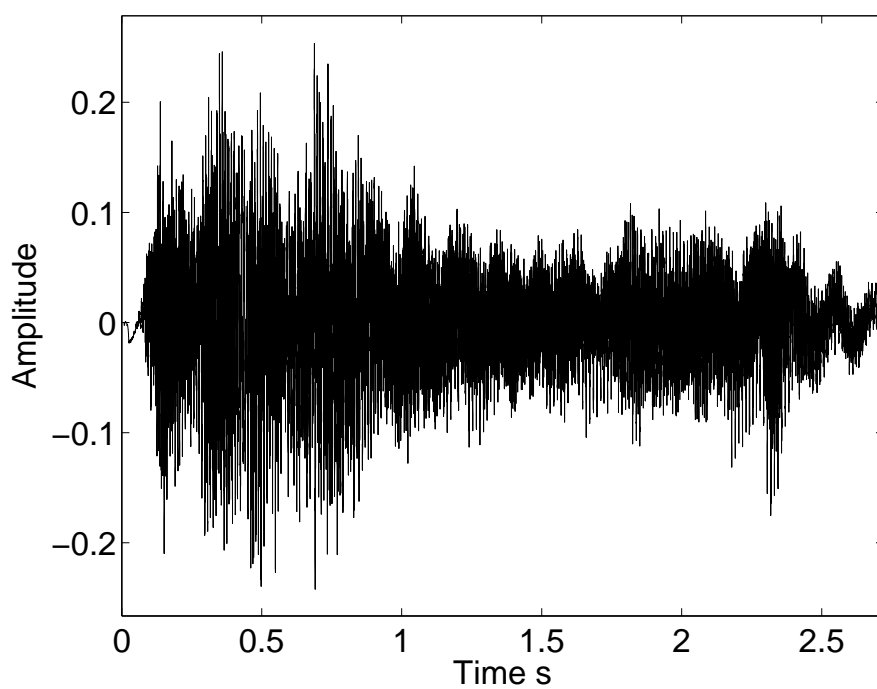


図 4.27: White-LIME による出力信号  $y(n)$  (44.1 kHz サンプルング：音楽信号)

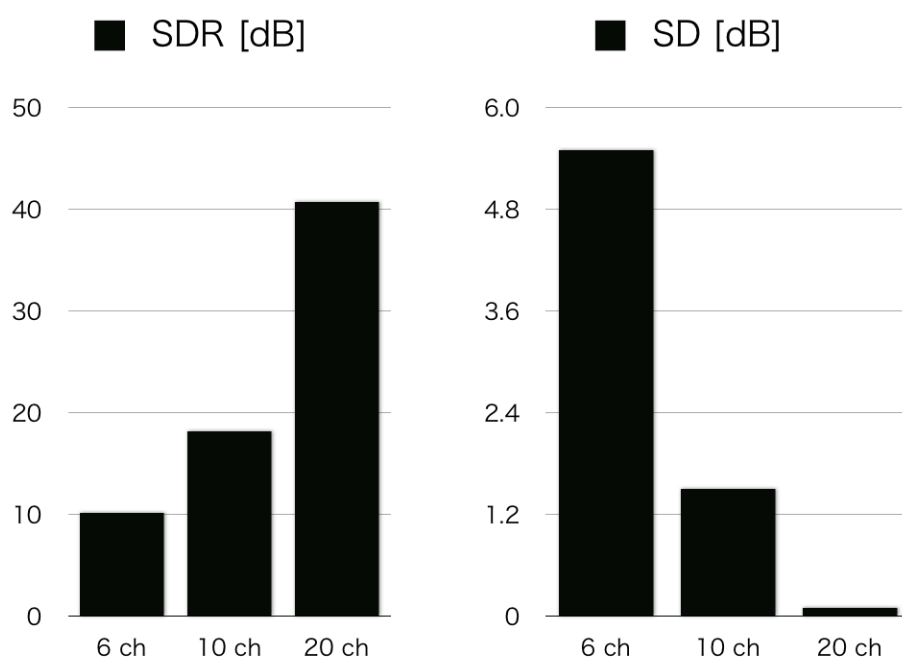


図 4.28: 原音が音楽信号の場合における White-LIME のチャンネル数の違いによる性能評価 (サンプルング周波数：44.1 kHz)

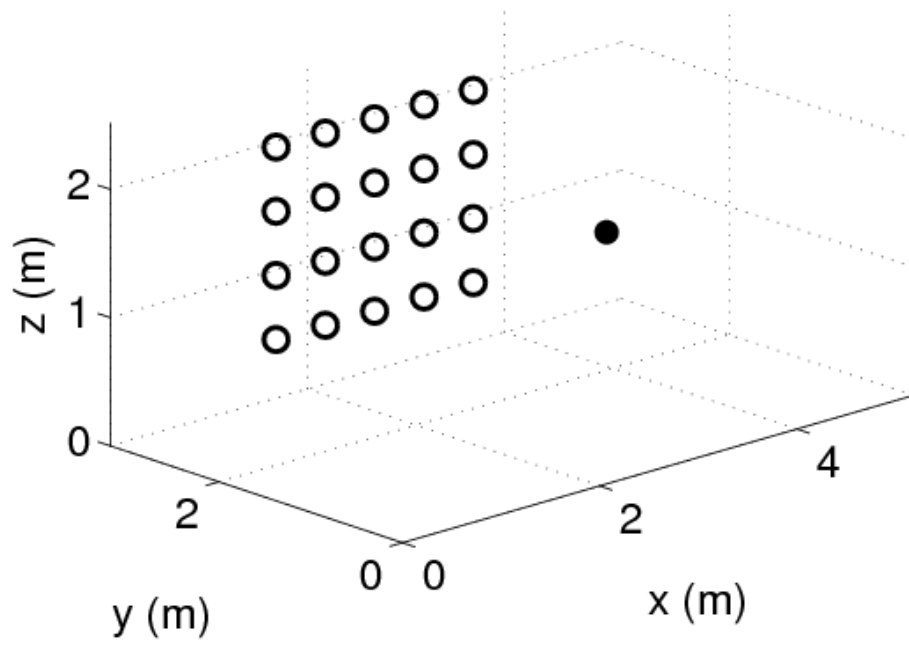


図 4.29: 20 ch を用いたマイクロホン配置 (スピーカ横側)

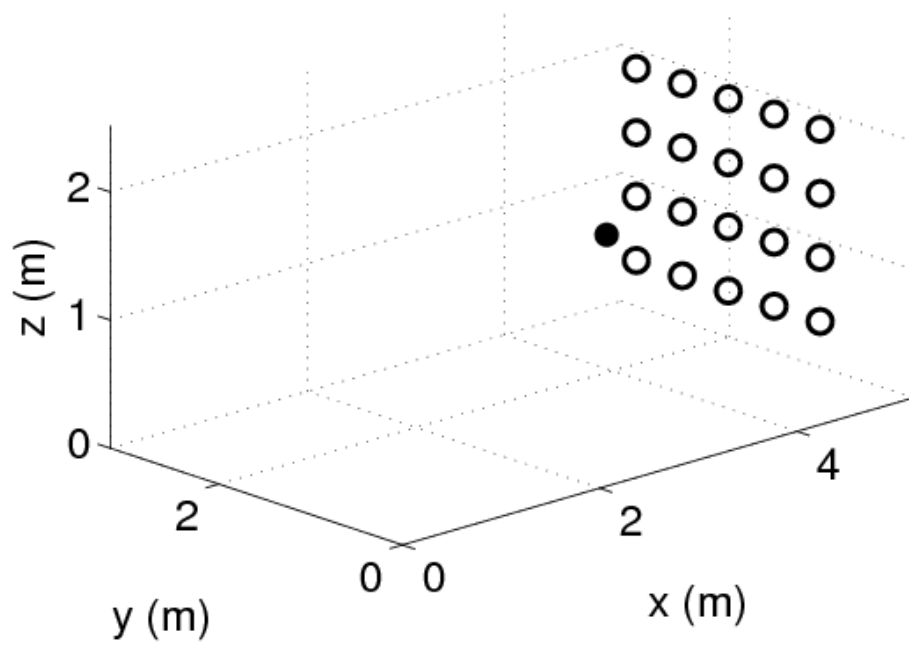


図 4.30: 20 ch を用いたマイクロホン配置 (スピーカ後方)



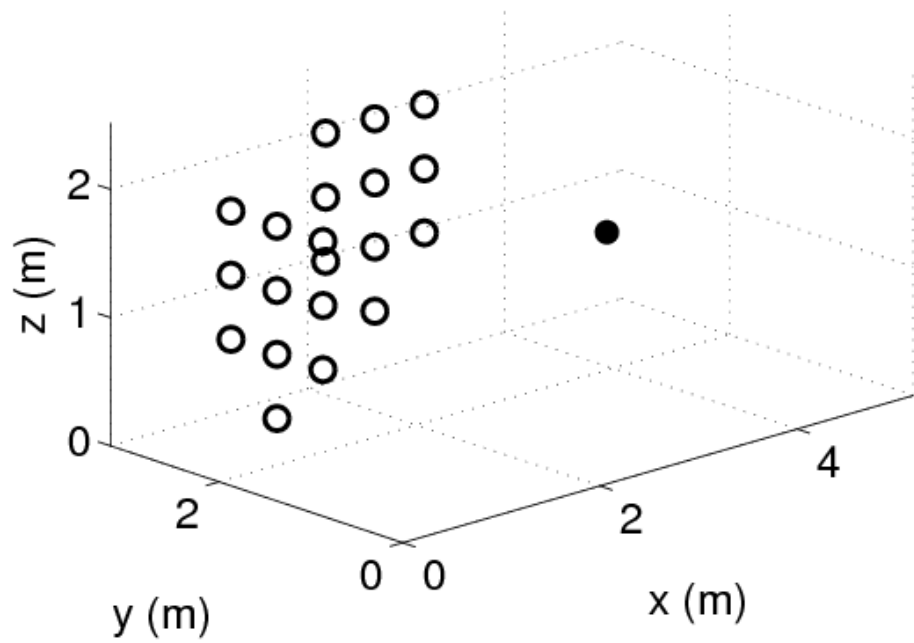


図 4.31: 20 ch を用いたマイクロホン配置 (スピーカ前方 10 ch , 横側 10 ch)

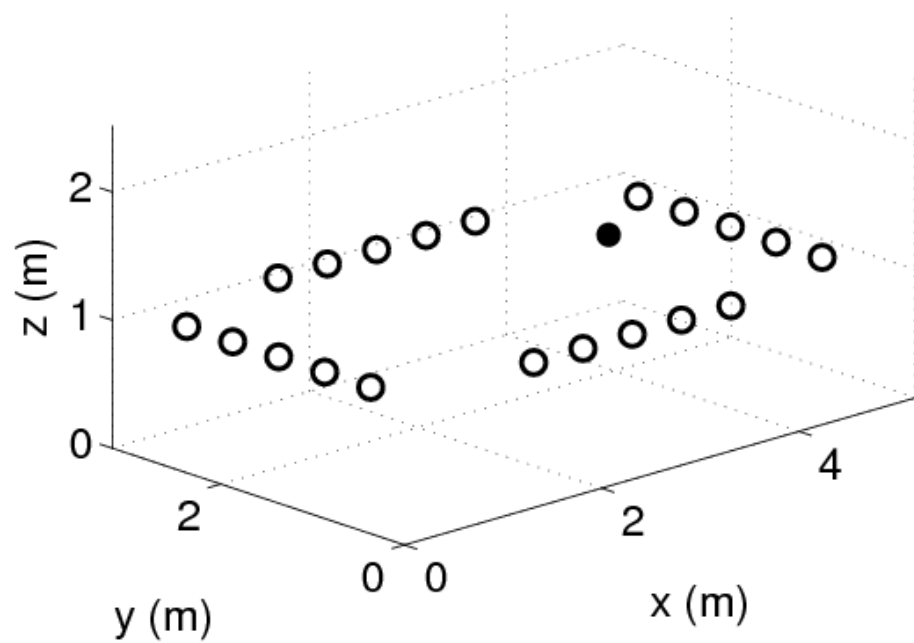


図 4.32: 20 ch を用いたマイクロホン配置 (スピーカ前後左右に各 5 ch ずつ)

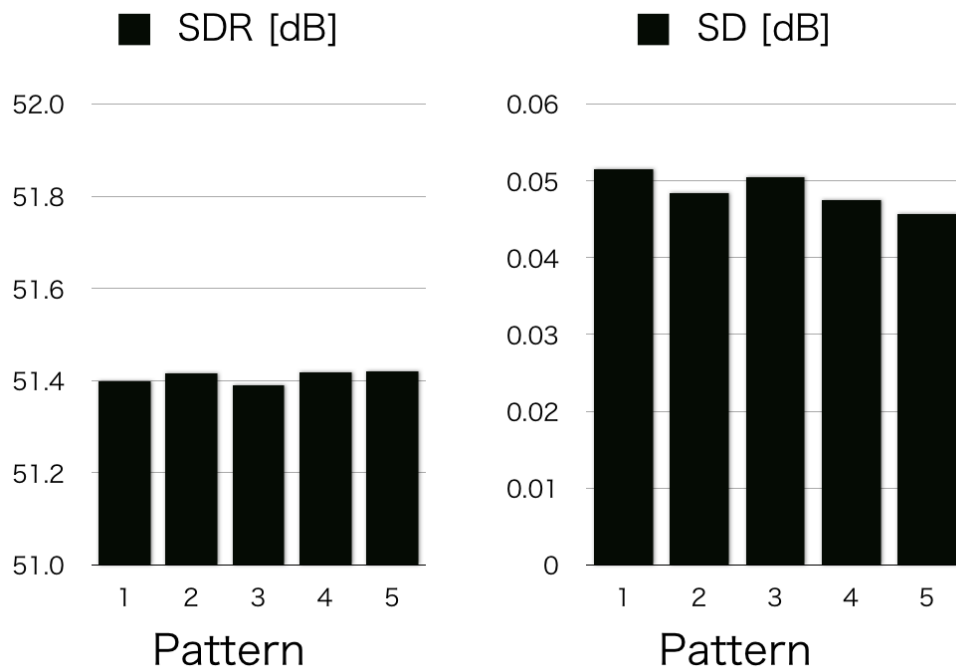


図 4.33: 原音が音声信号の場合における White-LIME のマイクロホン配置の違いによる性能評価 (サンプリング周波数: 24 kHz)

## 4.7 まとめ

本章では，室内残響が混入した観測信号から高精度に原音情報を抽出することが目的とした残響除去アルゴリズムの開発を行った．4.1 節において，本章の目的について述べた後，4.2 節では，多チャネルの観測信号から残響成分を完全に除去できる MINT 型逆フィルタの原理について述べた．4.3 節では，MINT 型逆フィルタを求めるために観測信号のみから室内インパルス応答を推定するブラインドシステム同定の原理および先行研究，問題点について概要した．4.4 節では，観測信号から MINT 型逆フィルタを直接推定するブラインド残響除去の原理および先行研究，問題点について述べた．4.5 節においては，ブラインドシステム同定とブラインド残響除去を比較し，本研究ではブラインド残響除去を用いたアルゴリズム開発を行うことを述べた．

以上の準備に基づき，4.6 節では従来の高精度残響除去アルゴリズム LIME が音声のようなスパース性のある音源の場合，長残響時や高サンプリング時には精度が劣化する問題を述べた．その上で，音源信号のスパース性を緩和する White-LIME を考案し，従来法の問題点を解決できることを述べた．また，White-LIME を用いた残響除去により，高サンプリング周波数の場合においても高精度に残響除去を行うことができることを示した．

残響除去に関する従来の研究はサンプリング周波数を 8 kHz や 16 kHz とした音声信

号に対する方法が主であった．それに対し White-LIME では，厳密な原音抽出のための高サンプリング周波数における残響除去を可能とした．その結果，提案した White-LIME を用いることにより，広帯域信号である音楽信号でも良好な原音抽出が可能であることを示した．

## 第5章 音源の放射指向特性の推定に関する検討

### 5.1 はじめに

本章では、音源の放射指向特性の抽出についての検討を行う。

実環境に存在する音源は振動パターンや音源自体に大きさがあるため、伝達関数も自由音場を仮定した点音源における球面波とは異なる。そのため、音源からの距離が等しくても、方向によって伝達関数が異なる。これを音源の放射指向特性と呼ぶ。

従来のマイクロホンアレイの研究の大半は、人間の顔やスピーカといった音源の正面から收音するという形をとっており、伝達関数には平面波や球面波を仮定している。したがって、音源の放射指向特性は考慮に入れていない。第1章でも述べた通り、文献 [28–30] では話者の放射指向特性の違いを人間がどれほど知覚できるかといった研究はなされているが [118–120]、放射特性そのものを推定しようとした研究はほとんど存在しない。音源の放射指向特性を考慮に入れた音源の位置と向きの同時推定という研究が中島らによって検討されているが [121]、音源の正面の音圧レベルが強く、後方に行くにしたがって弱くなるといった単純な仮定で音源の向きを推定する検討であるため、詳細な音源の放射指向抽出ではない。

包囲型マイクロホンは音源を囲むマイクロホン配置であるため、あらゆる方向からの直接音を取り込むことができる。この観測信号と第3章で得られた位置情報を利用することにより、音源位置からの放射指向特性は可能であると考えられる。しかし、これまで述べてきたように、室内の音場には反射音が存在するため、観測信号には音源の放射指向特性と反射音が混在して入力される。よって、音源の放射指向特性と反射音成分の分離が本章の課題である。

音源の放射指向特性と室内の残響特性分離を検討するために、5.2 節では、無響室での音源の放射指向特性のみの測定と室内における測定とを行い、両者の結果を比較する。5.3 節では、5.2 節での比較結果から、放射指向特性と室内残響の分離の指針および観測信号のみから各チャンネルのインパルス応答推定の必要性について述べる。5.4 節では、第4章で考案した White-LIME によって抽出した原音情報から室内インパルス応答を推定し、推定したインパルス応答の精度の評価を行う。5.5 節では、5.3 節での

指針と 5.4 節で得られたインパルス応答から，放射指向特性と室内残響の分離方法を提案する．無響室での測定結果と分離した応答を比較し，性能評価を行う．5.6 節で本章のまとめを述べる．

## 5.2 放射指向特性の測定

本節では，音源の放射指向特性と室内の残響特性の分離方法を検討するために，無響室における音源の放射指向特性の直接測定と，第 2 章で紹介した実験室において放射指向特性と室内残響とが混合した特性の測定を行う．音源には 1 ウェイのラウドスピーカを用いる．まずラウドスピーカとマイクロホン間のインパルス応答を無響室内および室内で測定する．前者の応答には反射音が存在しないと考えられるため，この応答を推定すべきラウドスピーカの放射指向特性と定義する．後者の応答は，ラウドスピーカの放射指向特性と室内残響特性の混合した特性となる．

### 5.2.1 無響室における放射指向特性の測定

ラウドスピーカの方向ごとの放射指向特性をインパルス応答として測定する．測定系を図 5.1，測定風景を図 5.2 に示す．図 5.2 のように，格子状のフレームの中心にマイクロホンを設置し，マイクロホンの正面方向 1.5 m の位置にスピーカを設置して測定を行う．また，使用した 1 ウェイスピーカは図 5.3 のように後方に穴が空いている．

測定はスピーカ正面を  $0^\circ$  とし，図 5.1 のようにスピーカを時計回りに  $15^\circ$  ずつ回転させ，計 13 方向の測定を行った．測定信号は第 4 章と同様，48 kHz サンプリングの TSP 信号 65536 ポイントを用いて，20 回の同期加算を行った観測信号である．5.4 節において 44.1 kHz サンプリングの音楽信号からインパルス応答の抽出を行うため，ここでのインパルス応答も 44.1 kHz サンプリングにダウンサンプリングして解析に用いる．測定結果の中から  $0^\circ$  方向， $90^\circ$  方向， $180^\circ$  方向のインパルス応答の時間波形と周波数特性を図 5.4 に示す．

### 5.2.2 室内における放射指向特性の測定

スピーカを部屋の中心に図 5.5 のように配置し，包囲型マイクロホンでのインパルス応答収録を行った．音源の高さは第 4 章同様  $z = 1.1$  m である．使用した TSP 信号は第 4 章および 5.2.1 項と同様である．本章ではスピーカと一番高さの近い  $z = 1.0$  m のマイクロホン 28 個を使用して放射指向特性の推定を行う．5.2.1 項同様，測定結果の中から  $0\text{ deg.}$ ， $90\text{ deg.}$ ， $180\text{ deg.}$  のインパルス応答の時間波形と周波数応答を図 5.6 に示す．

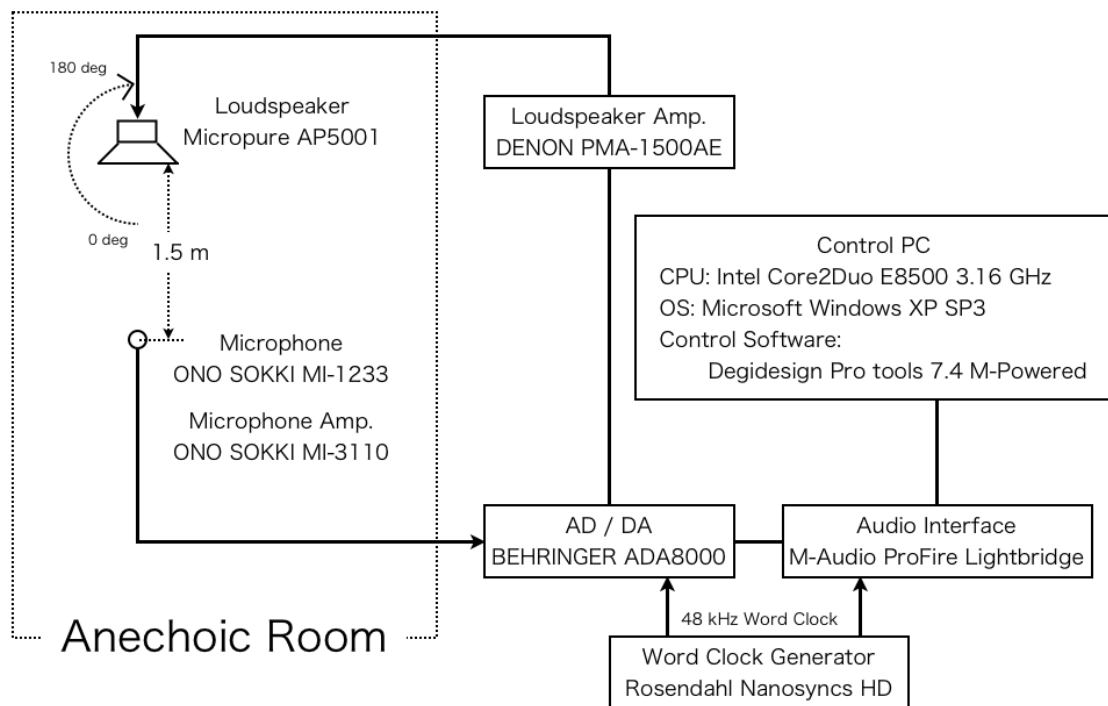


図 5.1: 無響室における放射指向特性の測定を表すシステム構成図

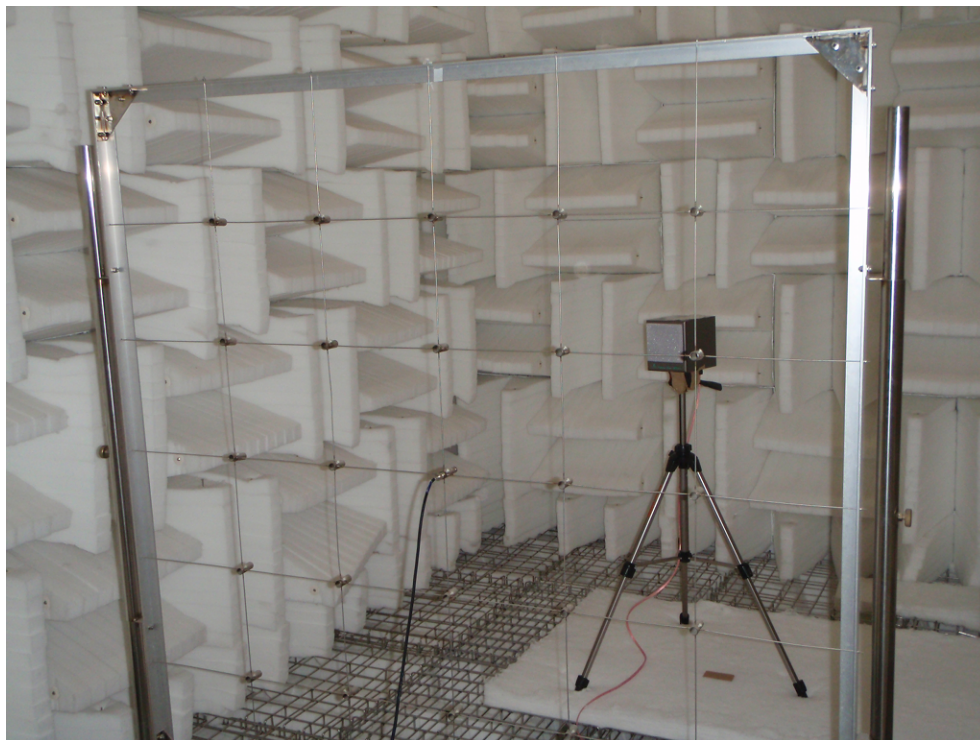


図 5.2: 無響室における測定の概観

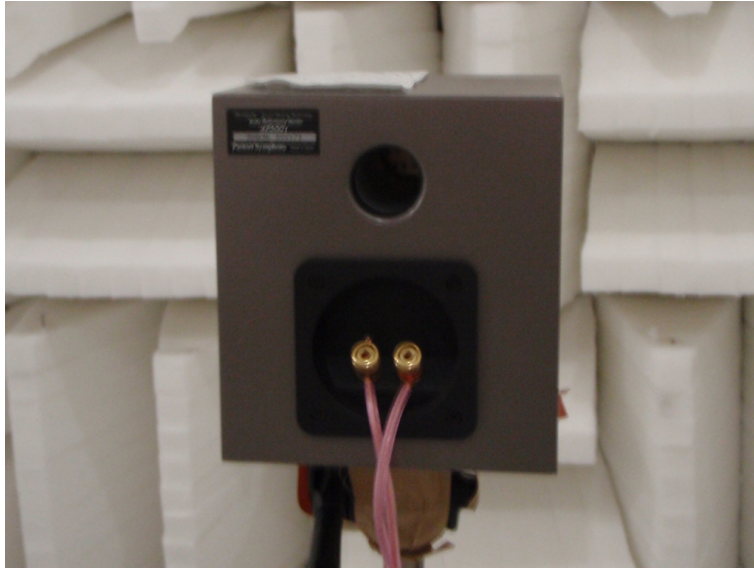


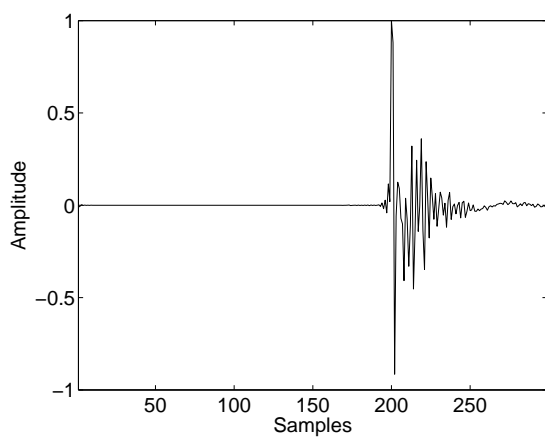
図 5.3: スピーカ後方の概観

### 5.2.3 測定結果の比較

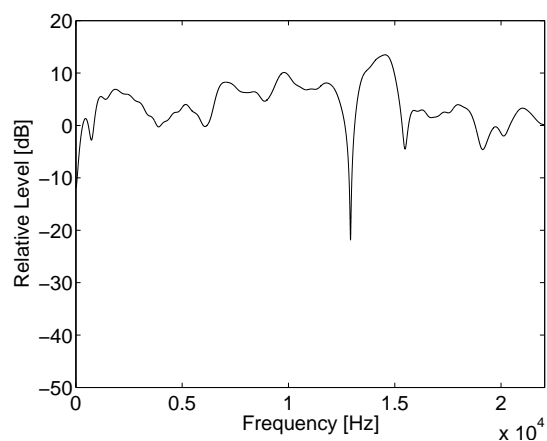
測定結果である図 5.4 と図 5.6 から，室内残響のない純粋な放射指向特性の伝達関数と放射指向特性と室内残響が混合された伝達関数の比較を行う．両者ともに共通する点は，前方の方が音圧レベルが高く，後方に行くにしたがって音圧レベルが下がる様子を確認できることである．これは，室内残響がない場合もある場合も，スピーカの前方にはスピーカの振動方向の音波が直接到来するため音圧レベルが高いが，側面に行くにしたがって振動方向から方向が変わるため，音圧レベルが下がることに起因していると考えられる．さらに後方に行くにしたがってスピーカの振動から伝わる音は少なくなり，スピーカユニットからの回折音が主応答になるため，音圧レベルがさらに下がる，回折は波長の長い低音に特有の物理現象であり高音では回折は起こりにくいため，後方に行くにしたがって高域の音圧レベルの方がより低下している様子が確認できる．

このように両者の全体的な傾向は似ているものの，方向ごと，周波数ごとの音圧レベル差は両者では異なる．これは，後者の方は室内残響特性も含んだ周波数特性であるために他ならない．

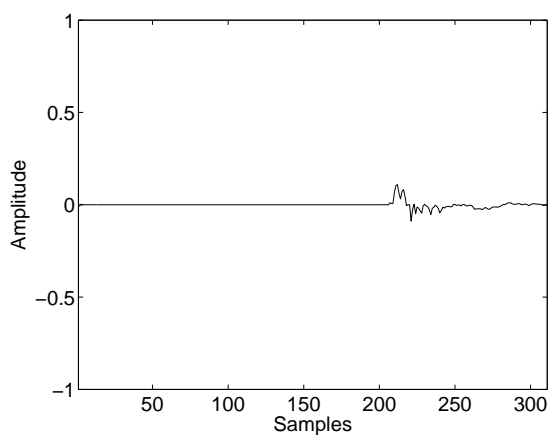
図 5.4(c) と図 5.6(c) に示した  $180^\circ$  のインパルス応答を比較する．室内での測定結果の，最初の音圧より少し遅れて到来した音圧の方が高いという現象は，スピーカからの回折音よりも，スピーカの前方に放射された音波の反射音の方が音圧レベルが高かったためであると考えられる．つまり，放射指向特性を含んだスピーカやユニットから直接届く音は最初に届き，その後には届く音は異なる方向に放射された音の反射音



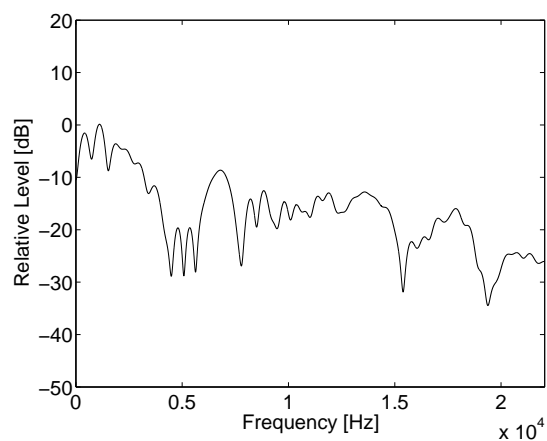
(a) 0 °方向のインパルス応答



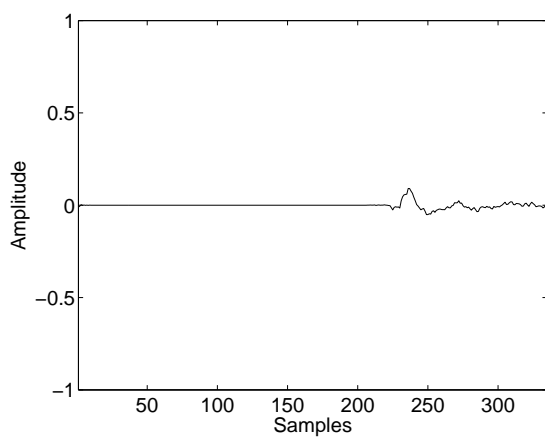
(d) 0 °方向の周波数特性



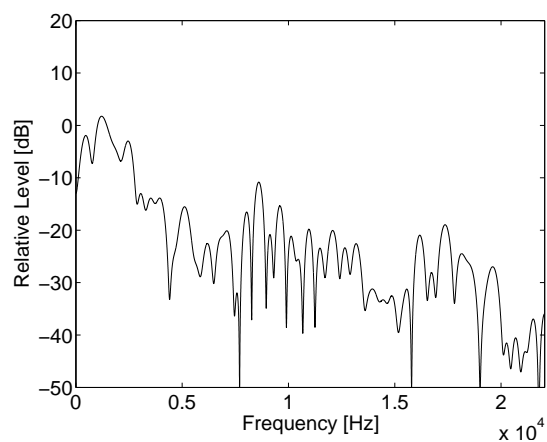
(b) 90 °方向のインパルス応答



(e) 90 °方向の周波数特性



(c) 180 °方向のインパルス応答



(f) 180 °方向の周波数特性

図 5.4: 無響室で測定した放射指向特性の測定結果



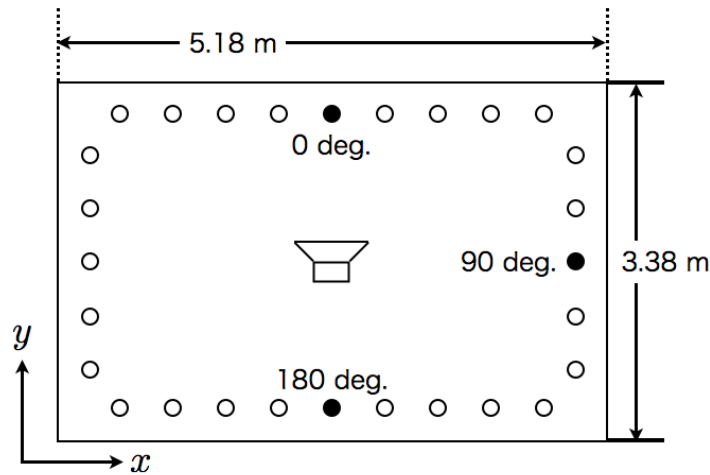


図 5.5: 室内におけるインパルス応答測定のマイクロホンの位置関係図 (パターン 1)  
 ○ : 測定結果を表示するマイクロホン , ● : 測定に使用するマイクロホン

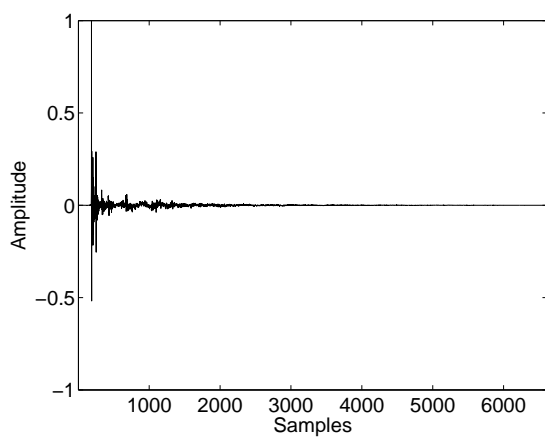
も届くことがわかる．つまり，様々な方向から様々な時間に届く反射音はそれぞれ別の放射指向特性を持っていることとなるため，反射音全体では様々な放射指向特性が混在した音となる．これは無響室で測定したインパルス応答にはない特徴である．

このように，インパルス応答，周波数特性で伝達関数の比較を行った結果，放射指向特性と残響特性の混合された伝達関数に対する残響成分の影響は非常に大きいことが確認できる．

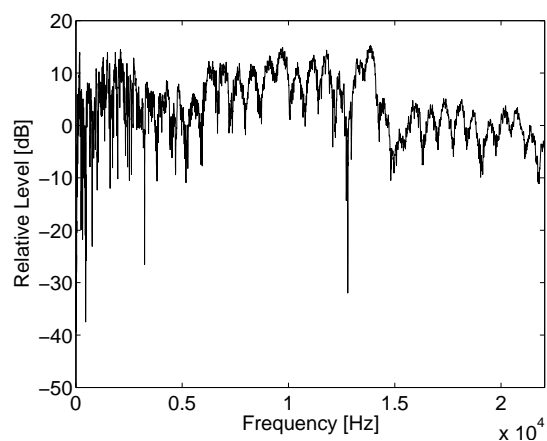
### 5.3 放射指向特性と室内残響分離の原理

音源に放射指向特性を持たない理想状態の場合と，音源に放射指向特性のある実環境の場合の伝達特性を図 5.7 と図 5.8 を用いて説明する．各観測点には音源からの直接音が最初に到来する．図 5.7 のように音源に放射指向特性がない点音源の場合は，インパルス応答の最初の 1 本目がこれに該当する．その後，直接音が壁に反射して各観測点に到来する初期反射，反射音がさらに壁に反射してから到来する後部残響の順に到来する．この順番は音源に放射指向特性がある場合にも同様である．両者の違いは，直接音の応答の有無である．放射指向特性を持つ音源の応答は図 5.8 のように点音源のように 1 本のインパルスではなく，図 5.6(a) ~ (c) の測定結果のように数十といった短いタップ数で構成されると考えられる．図 5.7 と図 5.8 はその違いを表す．

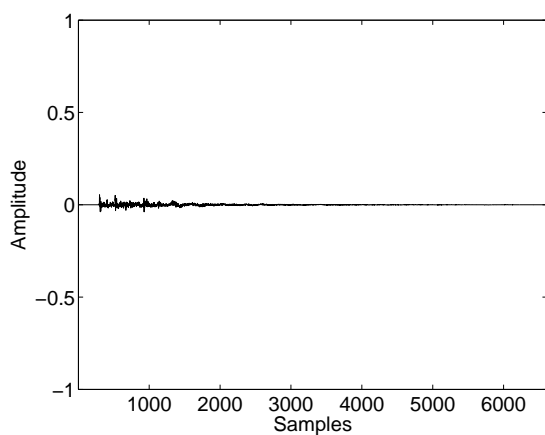
このように考えた場合，実環境での音源と観測点間のインパルス応答  $h(n)$  は，放射指向特性のインパルス応答  $h_{RA}(n)$  と室内残響特性のインパルス応答  $h_{RE}(n)$  は  $h(n) = h_{RA}(n) + h_{RE}(n)$  と時間軸での和であり，時間軸上ではほとんど重ならないと考えら



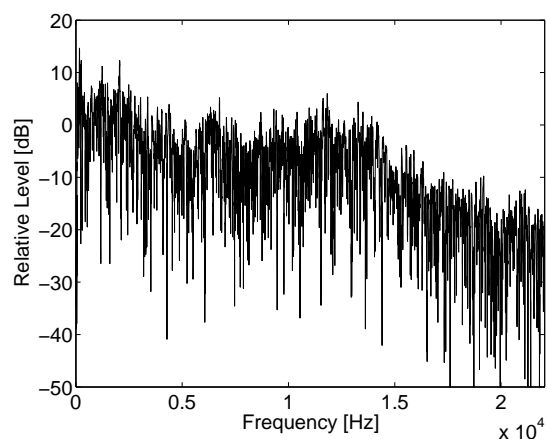
(a) 0 °方向のインパルス応答



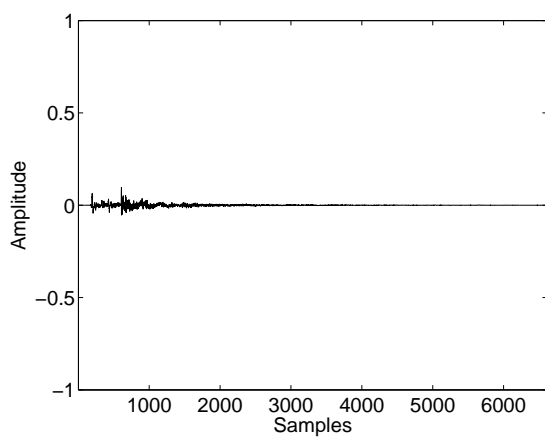
(d) 0 °方向の周波数特性



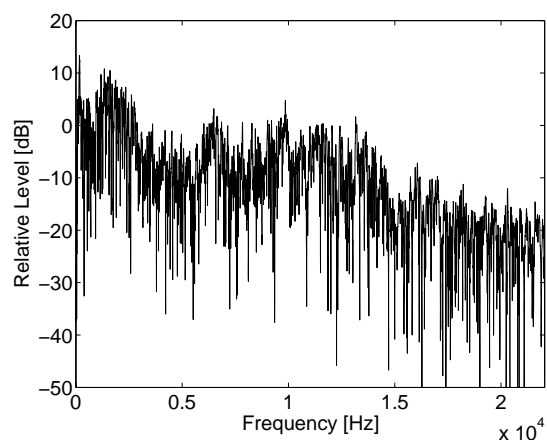
(b) 90 °方向のインパルス応答



(e) 90 °方向の周波数特性



(c) 180 °方向のインパルス応答



(f) 180 °方向の周波数特性

図 5.6: 室内で測定した放射指向特性と室内残響特性との混合した特性の測定結果

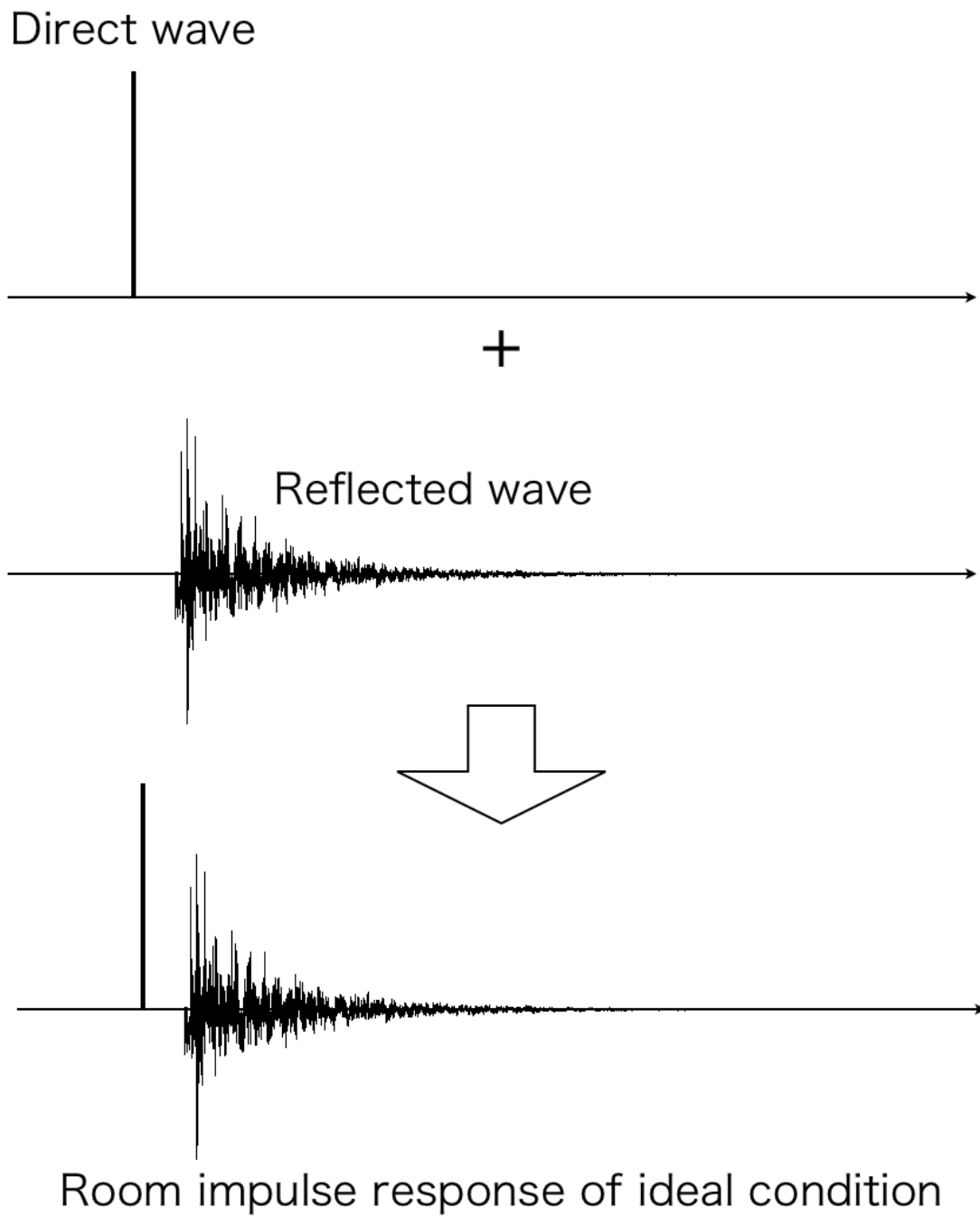
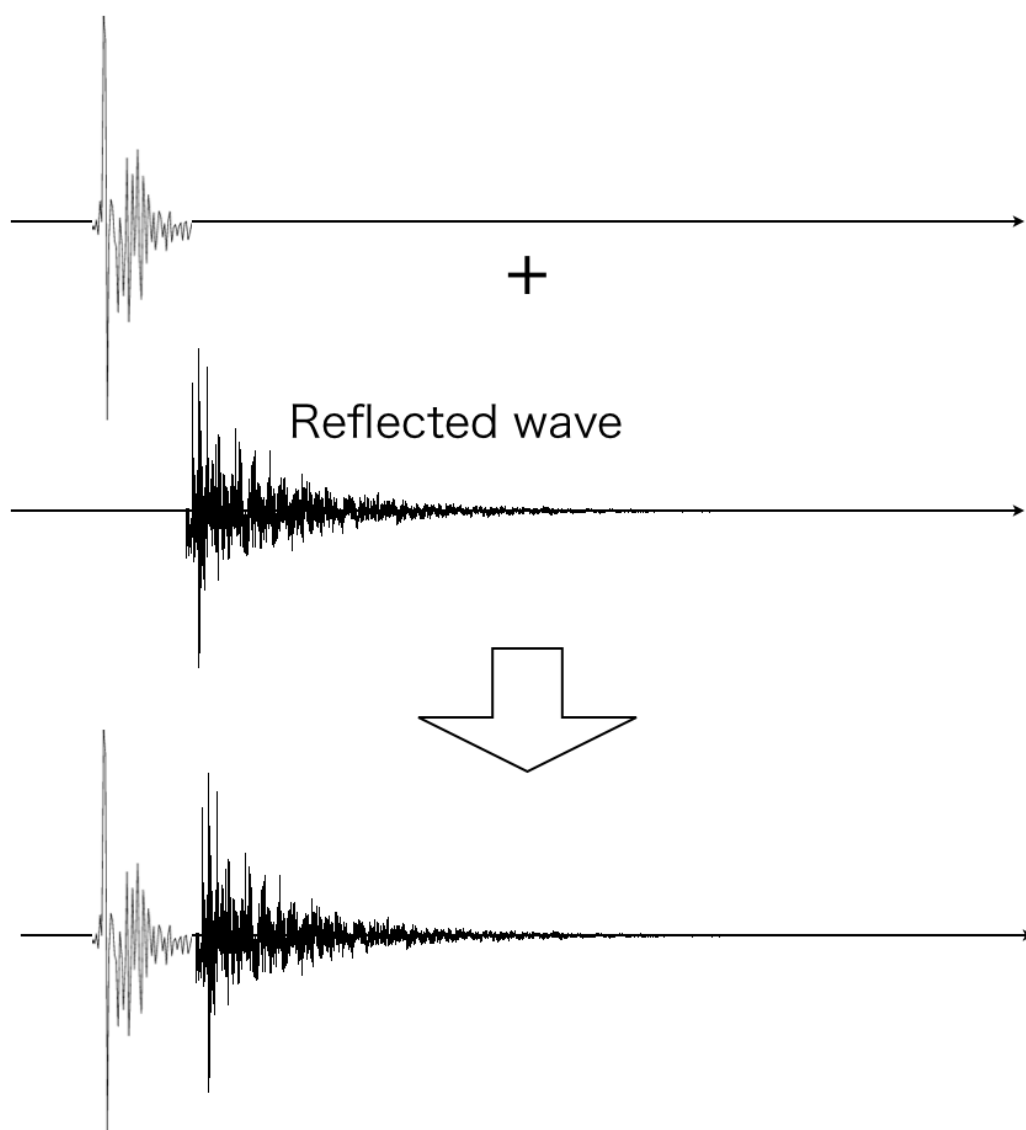


図 5.7: 音源が点音源である理想状態での室内インパルス応答

Radiation property



Room impulse response of real environment

図 5.8: 音源に放射指向特性がある実環境での室内インパルス応答

れる．

以下では，この2つの特性を分離する方針を考案する，放射指向特性はインパルス応答の最初の1本目が入力されてから数十タップの短い応答であると考えることができる．また，反射音は壁の反射を経路として伝搬される波であるため，直接音よりも時間が遅れて到来する．したがって，直接音と反射音には到来時間差が生じる，よって，直接音が到来してから初期反射音が到来するまでの時間の応答を切り出すことができれば，反射音成分を含まない放射指向特性のみを取り出すことが可能であると考ええる．観測点が壁から図5.9のように垂直に $d$ 離れて配置されていた場合， $c$ を音速とすると直接音を反射音の最短の到来時間差の最短 $t = 2d/c$ となる．よって，直接音の到来から時間 $t = 2d/c$ までのインパルス応答を切り出すことができれば，放射指向特性を抽出できると言える．ただし，そのためには観測点と壁面がある一定の距離だけ離れている必要がある，観測点を壁面が近接している場合は，切り出した放射指向特性は十分なタップ長が得られず近似解となる．

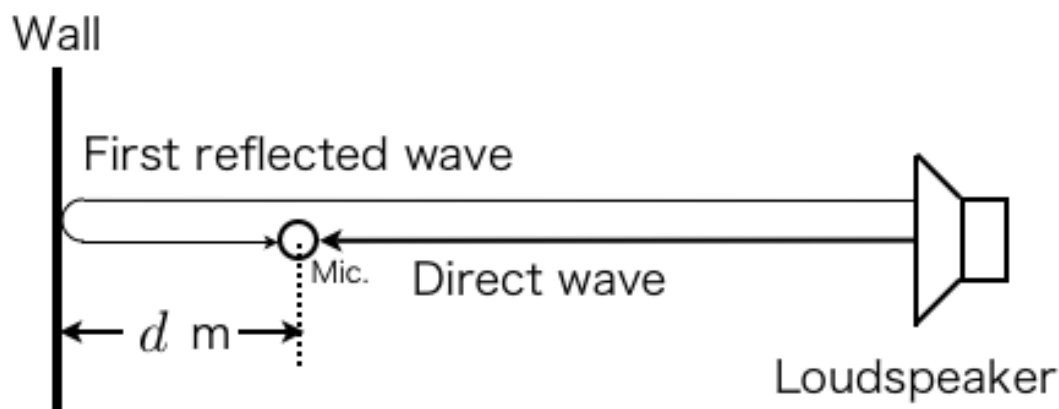


図 5.9: 直接音と初期反射音の伝搬経路差を表す概念図

このような方針で放射指向特性を抽出するためには，音源と各観測点間のインパルス応答を推定することが必要となる．第1章や第4章で述べた通り，インパルス応答の推定(=ブラインドシステム同定)はタップ数が未知であることやタップ長が非常に長いという理由から困難とされていた．しかし，第4章において考案した White-LIME によって原音の高精度抽出が可能となった，つまり，SIMO モデルにおける入力情報が明らかになったため，各チャンネルの観測信号を用いてインパルス応答の推定も可能となる．

よって，以下では，抽出した原音情報を用いたインパルス応答の抽出，さらには抽出したインパルス応答を用いた放射指向特性の抽出について検討する．

## 5.4 原音情報を用いたインパルス応答の抽出

### 5.4.1 原音情報を用いたインパルス応答の抽出の原理

第4章で考案した White-LIME によって観測信号から室内残響のみを取り除き，高精度に原音情報を抽出することが可能となった．よって本節では，原音情報と観測情報からインパルス応答の抽出について述べる．

LIME アルゴリズムでは線形予測に基づき予測誤差  $e(n)$  を求める．予測フィルタ長を  $L$  とした場合，時刻  $n$  の信号は  $n-1$  から  $n-L$  までの過去の  $L$  タップから計算される．したがって，LIME アルゴリズムにより計算される出力  $\hat{s}(n)$  は原音  $s(n)$  の  $L+1$  タップ目から出力される．

この関係を使い，インパルス応答  $\hat{h}_i(n)$  を推定する．推定方法の具体的な流れを図 5.10 に示す．原音  $s(n)$  のタップ長を  $U$  とする．まず，LIME の出力結果  $\hat{s}(n)$  は， $s(n)$  の  $L+1$  タップ目から  $U$  タップまで抽出され，White-LIME 過程での原音の平均スペクトルや Pre-whitening フィルタの逆畳み込みによって残りはすべて 0 として出力される．

この出力結果  $\hat{s}(n)$  を原音と見なし各観測信号  $x_i(n)$  から逆畳み込みすることにより，インパルス応答  $\hat{h}_i(n)$  を抽出する．具体的には，図 5.10 のように  $\hat{s}(n)$  の最初の  $L$  タップに 0 を挿入し，各観測信号  $x_i(n)$  と同じタップ数で離散フーリエ変換 (DFT) する (FFT の場合は  $x_i(n)$  と同じタップ数より大きい 2 のべき乗数のタップを用いる)．各観測信号  $x_i(n)$  も同様に DFT する．得られた周波数領域の信号  $X_i(f)$  と  $\hat{S}(f)$  から，

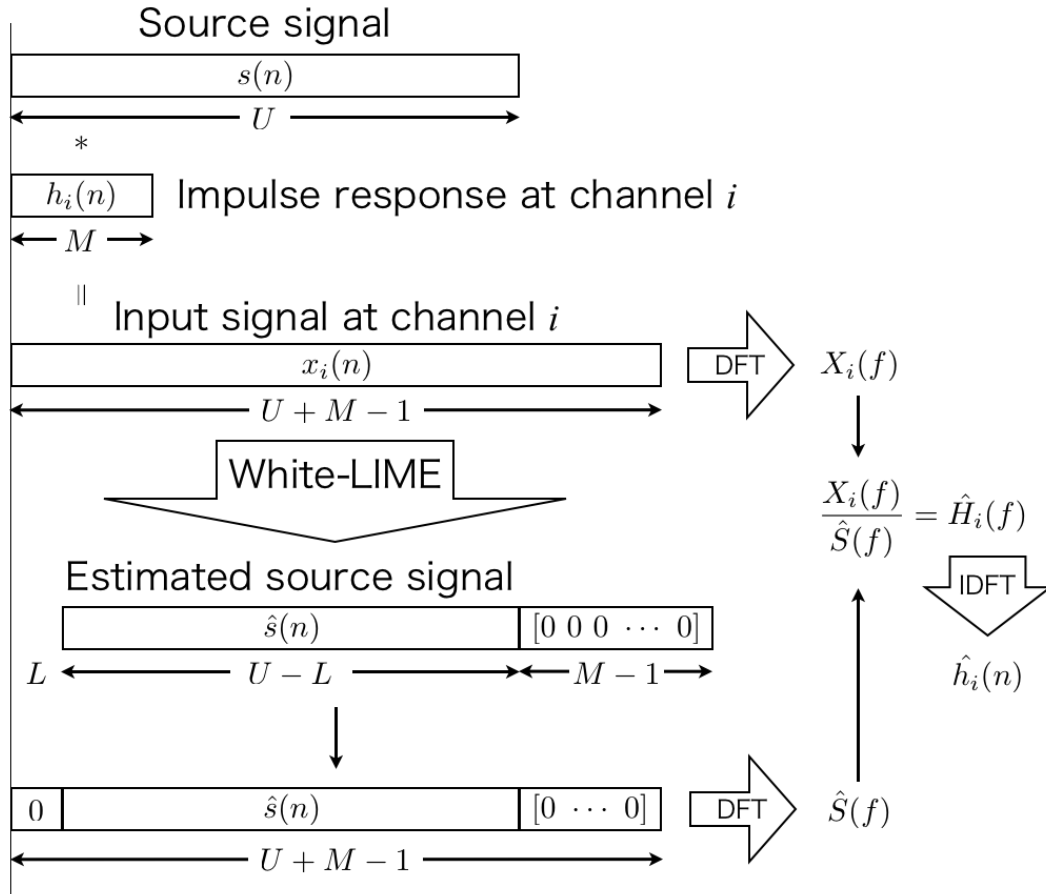
$$\hat{H}_i(f) = \frac{X_i(f)}{\hat{S}(f)} \quad (5.1)$$

を計算することにより  $\hat{H}_i(f)$  が得られ，離散逆フーリエ変換 (IDFT) して時間領域に戻すことによってインパルス応答  $\hat{h}_i(n)$  を抽出できる．

LIME アルゴリズムは線形予測に基づく方法のため，原音の最初の  $L$  タップは抽出できない．しかし，式 (4.77) のようにチャネル数  $P$  を増やすことによって  $L$  を短くすることができる．さらに，観測信号が長い場合， $L$  は全体のタップ数から見てもかなり小さな値とできるため， $L$  タップの 0 挿入は微小な誤差で済むこととなる．

### 5.4.2 インパルス応答抽出の性能評価

小節 5.2.2 で測定した 28 チャネルのインパルス応答を 44.1 kHz にダウンサンプリングしたインパルス応答 6615 タップと第4章で用いた 44.1 kHz の音楽信号約 2.7 s (120000 タップ) を畳み込んだ信号を観測信号  $x_i(n)$   $i = 28$  とし，White-LIME によっ



$\hat{h}_i(n)$  : Estimated impulse response at channel  $i$

図 5.10: 観測信号  $x_i(n)$  と抽出原音  $\hat{s}(n)$  の逆畳み込みによるインパルス応答  $\hat{h}_i(n)$  の抽出

て残響除去を行い出力信号  $\hat{s}(n)$  を得る．第 4 章でも用いた評価指標 SD を原音  $s(n)$  と出力  $\hat{s}(n)$  で計算したところ，57.3 dB と非常に高い値が得られた．これは，高精度に原音を抽出することができたことを示している．5.1 節でも述べた通り，話者の放射指向特性は注目がなされており，話者の放射指向特性を推定するためには，音声信号を用いることが想定される．しかし，本章においては放射指向特性の基本的な性能を確認したいと考えているため，音声よりもより広帯域な音楽信号を音源として用いることとした．

LIME アルゴリズムでは，各信号に共通して畳み込まれた要素が出力として残るという性質がある．LIME アルゴリズムでは，残響と同時に音源と各観測点の放射指向特性も完全に除去している．これは，放射指向特性が各観測点で異なることと，同じ観測点であっても残響成分の放射指向特性は様々であるためである．

高精度に抽出した出力  $\hat{s}(n)$  と各観測信号  $x_i(n)$  から式 (5.1) を計算し，IFFT することによって音源と各観測点間のインパルス応答  $\hat{h}_i(n)$  を抽出した．元の長さ 6615 タップにそろえて相関係数を計算したところ，どのチャンネルもほぼ 1 であった．先ほど用いた評価指標 SD を  $h_i(n)$  と  $\hat{h}_i(n)$  で計算し，全チャンネルの平均値をとったところ，62.6 dB と極めて高い値であった．White-LIME から得られた原音を用いることによって，音源と観測点間のインパルス応答をほぼ正確に算出できることが確認できる．

## 5.5 インパルス応答からの放射指向特性抽出

### 5.5.1 インパルス応答切り出しによる放射指向特性の抽出

5.3 節で示した方針に基づき，5.4.2 項で抽出したインパルス応答の切り出しによって放射指向特性の抽出を試みる．

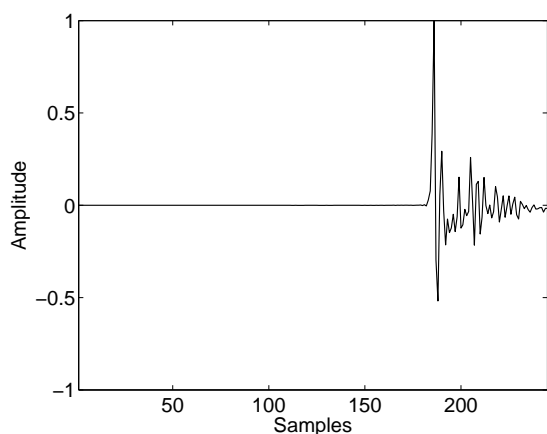
測定した環境では，図 2.3 に示した通り，壁面とマイクロホン間の距離は 30 cm である．したがって，44.1 kHz サンプリングのインパルス応答の場合，音速  $c = 340$  m/s とすると， $44100 \times 2 \times 0.6 / 340 \approx 78$  タップとなる．

無響室で測定した応答の初期応答から 78 タップを切り出した応答の中から  $0^\circ$ ， $90^\circ$ ， $180^\circ$  のインパルス応答の時間波形，周波数応答，位相特性を図 5.11 に示す．

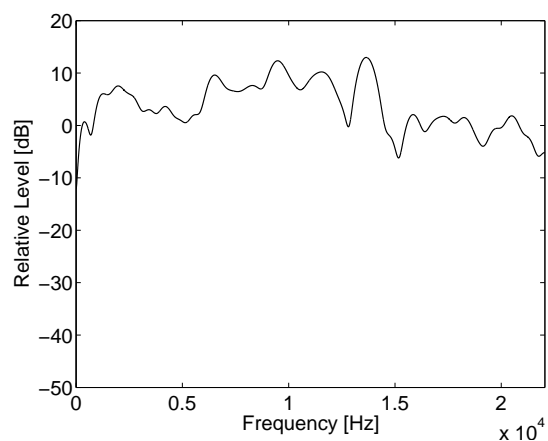
### 5.5.2 抽出した放射指向特性の性能評価

無響室における測定結果 (図 5.4)，室内での測定結果 (図 5.6)，抽出結果 (図 5.11) の 3 種類の応答を比較する．

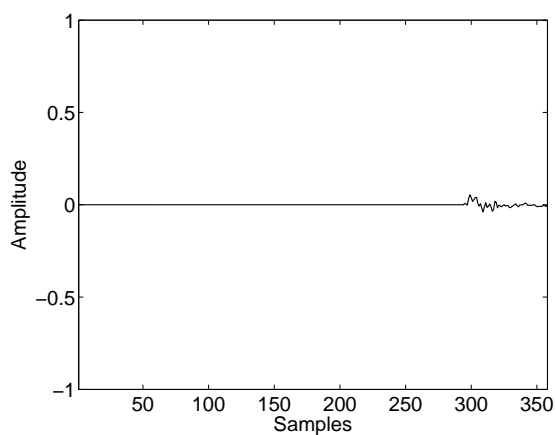




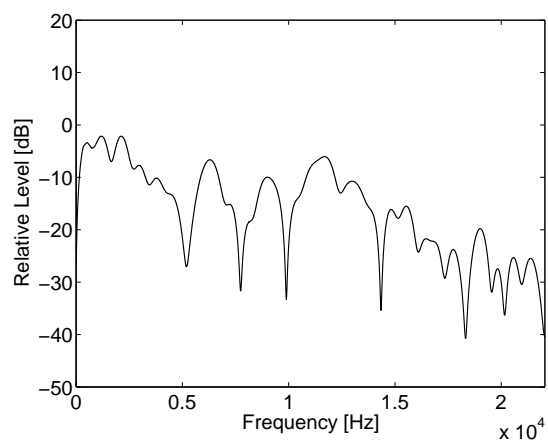
(a) 0 °方向のインパルス応答



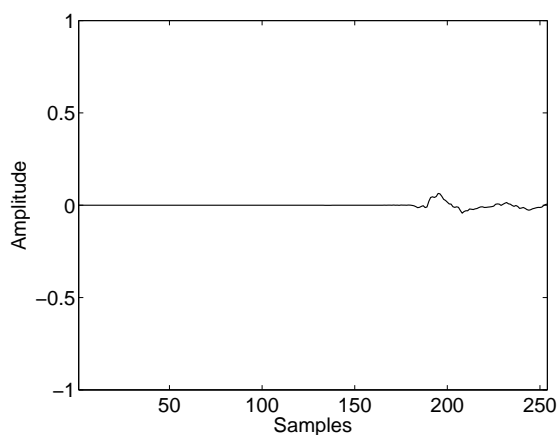
(d) 0 °方向の周波数特性



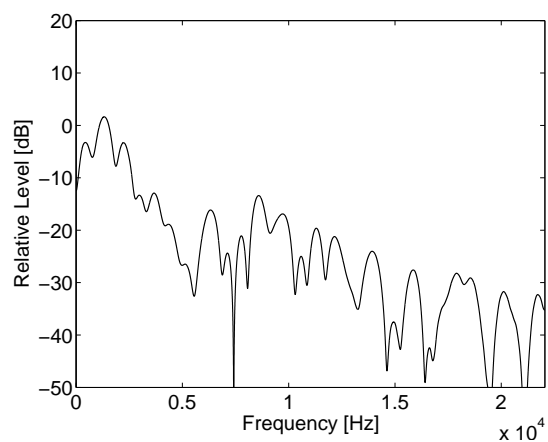
(b) 90 °方向のインパルス応答



(e) 90 °方向の周波数特性



(c) 180 °方向のインパルス応答



(f) 180 °方向の周波数特性

図 5.11: 観測信号から抽出したインパルス応答を切り出して得られる放射指向特性

まず， $0^\circ$ ， $90^\circ$ ， $180^\circ$ における周波数特性であるが，無響室での応答と室内での応答は全く異なる傾向であるのに対して，抽出した応答ではほぼ同様の傾向を示している．

実環境では室内の中心に音源がある場合だけではなく，音源は様々な位置に様々な向きで存在する．そこで，スピーカの方を変えたパターン 2 (図 5.12)，スピーカの位置を変えたパターン 3 (図 5.13)，スピーカの向きと位置を変えたパターン 4 (図 5.14) についても同様の測定を行った．

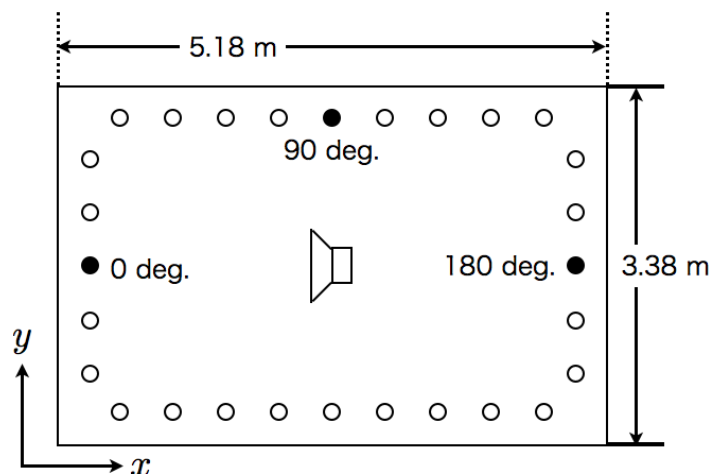


図 5.12: 室内におけるインパルス応答測定のマイクロホンの位置関係図 (パターン 2)  
○：測定結果を表示するマイクロホン，●：測定に使用するマイクロホン

その上で，無響室における測定結果，室内における測定結果 (パターン 1)，抽出結果 (パターン 1～4) の計 6 種類の比較を方向ごとの  $1/3$  オクターブバンド解析によって行う．無響室での測定結果は 25 方向の，それ以外では 28 方向の応答を  $1/3$  オクターブ分析し，各中心周波数ごとの方向による音圧の違い  $125\text{ Hz} \sim 16\text{ kHz}$  の 10 パターンを図 5.15～5.24 に示す．ここで，室内における測定結果と抽出結果 (パターン 1～4) については，音源から各マイクロホンまでの距離が異なるため，距離補正が必要となる．そこで，切り出した放射指向特性のインパルス応答 (チャンネル  $i$ ) を  $h'_{Di}(n)$  とし，音源とチャンネル  $i$  までの距離を  $r_i$  とすると，距離補正後のインパルス応答  $h_{Di}$  は，

$$h_{Di} = r_i h'_{Di}(n) \quad (5.2)$$

となる．距離の算出は，第 3 章で考案した RAP-MUSIC から得られる音源位置情報と各観測点の位置情報から推定が可能である．室内における測定結果と抽出結果 (パターン 1～4) については距離補正した応答を用いている．

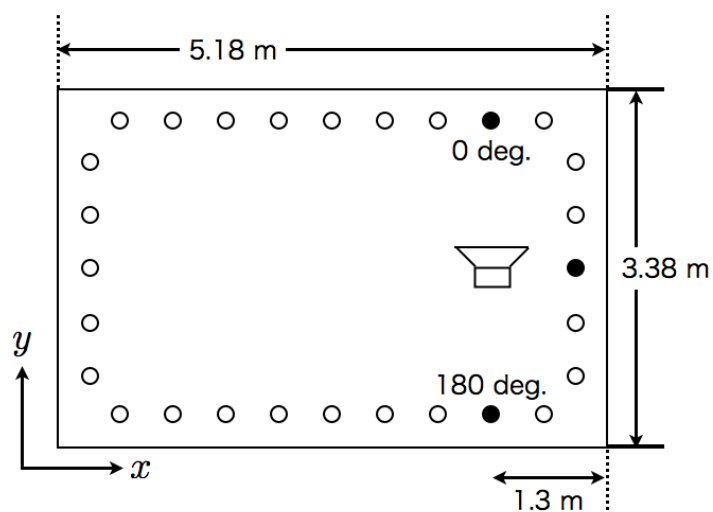


図 5.13: 室内におけるインパルス応答測定のマイクロホンの位置関係図 (パターン 3)  
 ○ : 測定結果を表示するマイクロホン , ● : 測定に使用するマイクロホン

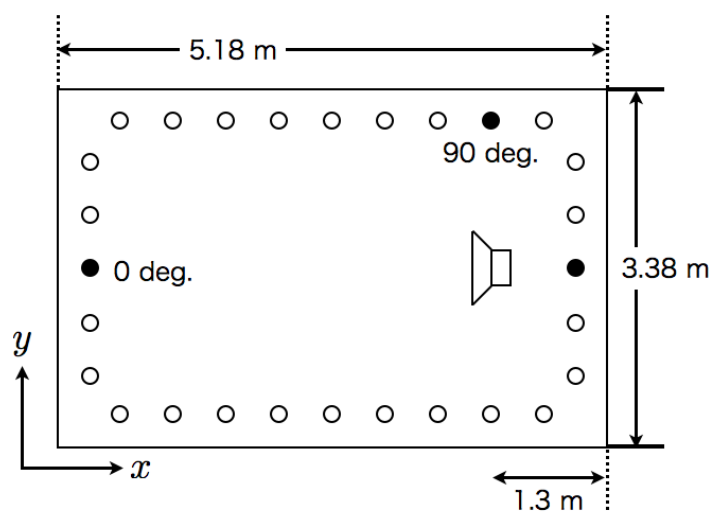


図 5.14: 室内におけるインパルス応答測定のマイクロホンの位置関係図 (パターン 4)  
 ○ : 測定結果を表示するマイクロホン , ● : 測定に使用するマイクロホン

方向ごとの 1/3 オクターブバンド解析の結果では，音源からマイクロホンまでの距離が異なる位置で計測した結果にも関わらず，500 Hz 以降では無響室での応答とほぼ同様の傾向であることが確認できる．

また，方向ごとの 1/3 オクターブバンド解析の結果をパターンマッチングを用いた最短距離法に基づく類似度  $S$  と余弦尺度  $\cos\theta$  を用いて評価する．最短距離法に基づく類似度はベクトル間の距離の評価指標である．余弦尺度はベクトル間の向きの評価指標であり，ベクトル間の相関係数に対応する．無響室での測定結果の方向ごとの音圧分布のパターンベクトルを  $P_1(\tau)$ ，室内での測定結果の方向ごとの音圧分布のパターンベクトルを  $P_2(\tau)$ ，抽出結果の方向ごとの音圧分布のパターンベクトルを  $P_3(\tau)$  とする．無響室での測定結果と室内での測定結果の最短距離法に基づく類似度を  $S_{1,2}$ ，無響室での測定結果と抽出結果の最短距離法に基づく類似度を  $S_{1,3}$  とすると，

$$S_{1,2} = 1 - \frac{|P_1(\tau) - P_2(\tau)|}{|P_1(\tau)|} \quad (5.3)$$

$$S_{1,3} = 1 - \frac{|P_1(\tau) - P_3(\tau)|}{|P_1(\tau)|} \quad (5.4)$$

により算出する．また，無響室での測定結果と室内での測定結果の余弦尺度  $\cos\theta_{1,2}$ ，無響室での測定結果と抽出結果の余弦尺度を  $\cos\theta_{1,3}$  とすると，

$$\cos\theta_{1,2} = \frac{P_1(\tau) \cdot P_2(\tau)}{|P_1(\tau)| |P_2(\tau)|} \quad (5.5)$$

$$\cos\theta_{1,3} = \frac{P_1(\tau) \cdot P_3(\tau)}{|P_1(\tau)| |P_3(\tau)|} \quad (5.6)$$

となる．最短距離法に基づく類似度は音圧分布の方向ベクトルの距離を表し，余弦尺度は音圧分布の方向ごとにプロットしたグラフの概形の類似度を表す．

まずはパターン 1 の結果について考察する．各バンドごとの最短距離法に基づく類似度と余弦尺度を図 5.25 に示す．無響室での測定結果と室内での測定結果を見ると，低域は比較的 1 に近い値を取っているのに対し，高域では最短距離法に基づく類似度では 0.5 付近と小さくなっている．低域成分は波長が長く，音波自体の指向性は低いいため，残響特性が混在している場合も，無響室での測定結果と同様の傾向を示す．一方，高域成分は波長が短く，音波自体の指向性も鋭いため，残響特性が混在した場合は，無響室での測定結果とは音圧分布が大きく異なる．図 5.25 の結果はこの状態を表していると言える．

それに対して，無響室での測定結果と抽出結果の類似度は最短距離法に基づく類似度は室内での測定結果との類似度よりも全て高い値となっている．特に，室内での測

定では 0.5 付近であった高域でも抽出結果の場合 0.9 付近の値である．つまり，インパルス応答を切り出すことによって無響室での放射指向特性に近づくことがわかる．

パターン 2, 3 の結果 (図 5.26, 図 5.27) もパターン 1 と同様の傾向を示している．

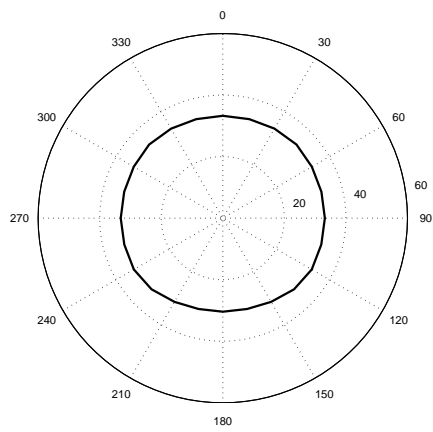
パターン 4 の結果 (図 5.28) は，室内での測定結果も類似度が高い．これは，スピーカ後方側と観測点との距離が近いこと，スピーカ後方の放射指向特性を他のパターンよりも高い SN 比 (S: 放射指向特性, N: 反射音) で観測しているためであると考えられる．しかしパターン 4 の結果も，他のパターンと同様，抽出した放射指向特性の方が全帯域において類似度が大きいことに変わりはない．

以上の結果から，室内インパルスの初期応答から，初期反射までの時間のインパルスを切り出すことができれば，音源の放射特性を抽出できることがわかる．

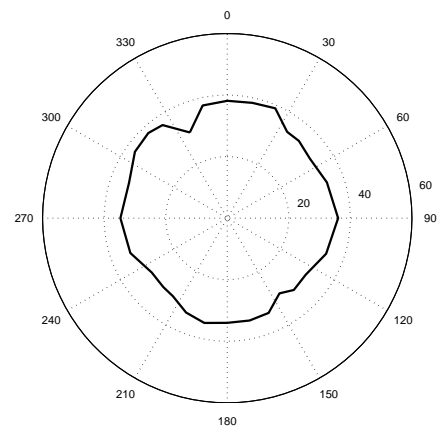
## 5.6 まとめ

本章においては，反射音の存在する室内における観測信号のみから音源の放射指向特性の抽出についての検討を行った．

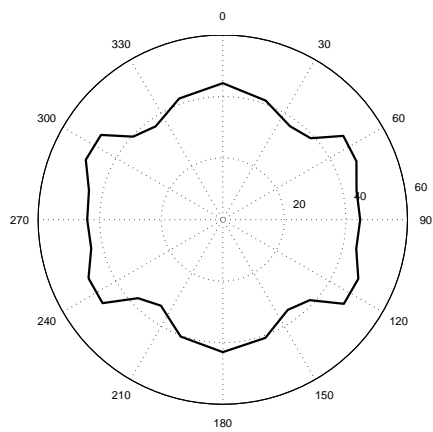
実環境の音源の放射指向特性を抽出するという研究はこれまでに前例がない．5.2 節において放射指向特性そのものの測定および室内での反射音を含んだ測定を行い，両者の違いを比較した．5.3 節では 5.2 節の考察をもとに，放射指向特性と室内残響成分の分離に関する指針を提示した．5.4 節では，放射指向特性と室内残響成分の分離のためのインパルス応答の抽出に関する検討を行った．第 4 章にて考案した White-LIME によって得られる原音情報を用いることによって，ほぼ正確なインパルス応答が抽出できることをミュレーションによる性能評価によって明らかにした．5.5 節においては，抽出したインパルス応答の初期反射までの時間を切り抜いた応答を用いることにより，反射音の影響を低減し，音源の放射指向特性をほぼ再現できることを性能評価によって確認した．得られた放射指向特性と Pre-Whitening LIME によって得られる原音信号を畳み込むことによって音源の方向ごとの放射指向特性を再現することができると考える．また，残りのインパルス応答は観測点における反射音成分として抽出できたとも言え，今後の活用が期待できる．



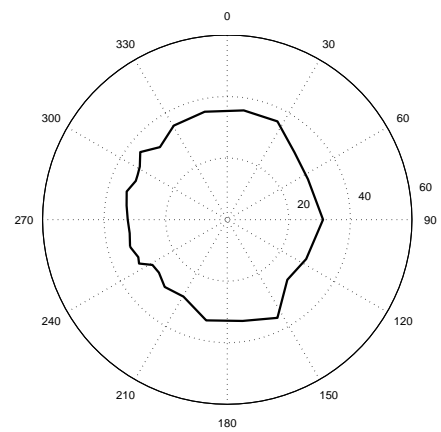
(a) 無響室での測定結果



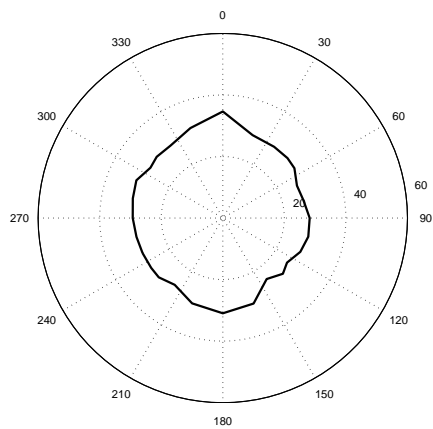
(d) パターン 2 の抽出結果



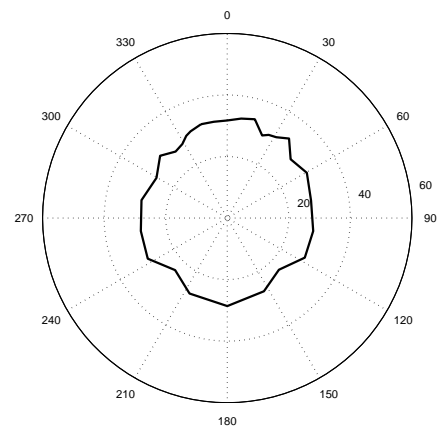
(b) 室内での測定結果 (パターン 1)



(e) パターン 3 の抽出結果

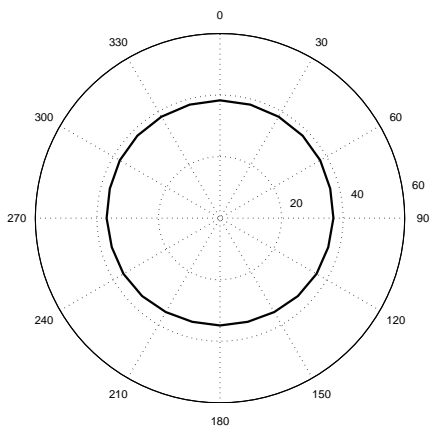


(c) パターン 1 の抽出結果

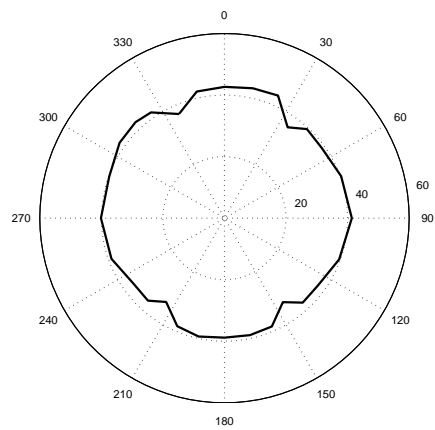


(f) パターン 4 の抽出結果

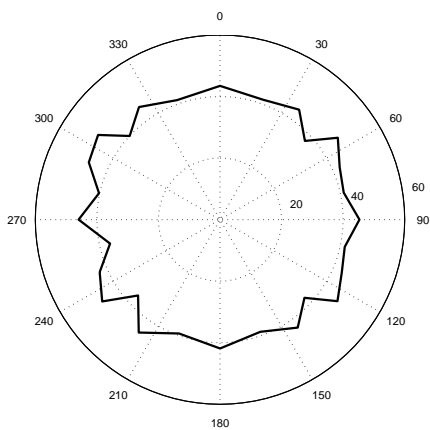
図 5.15: 方向ごとの放射指向特性の違いによるスピーカの音圧分布 (中心周波数:125 Hz)



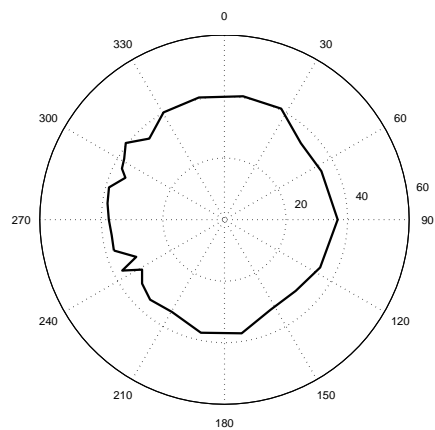
(a) 無響室での測定結果



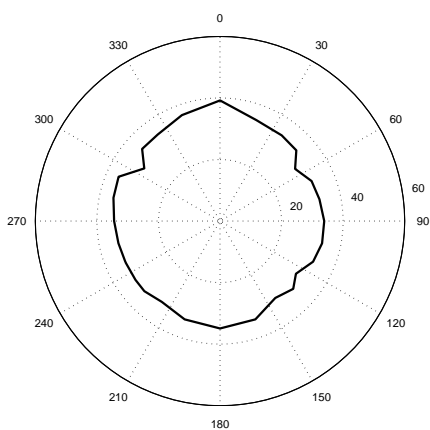
(d) パターン 2 の抽出結果



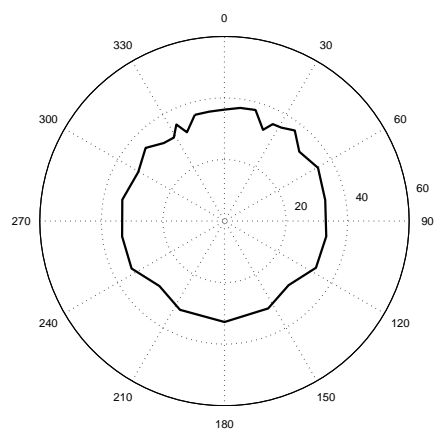
(b) 室内での測定結果 (パターン 1)



(e) パターン 3 の抽出結果

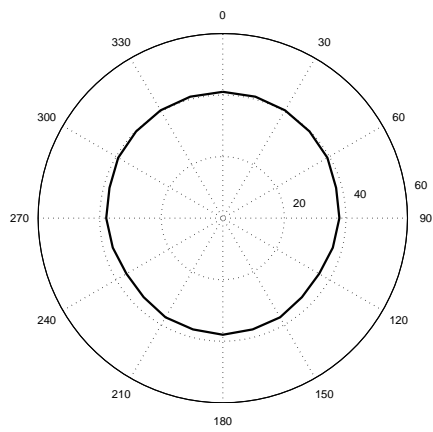


(c) パターン 1 の抽出結果

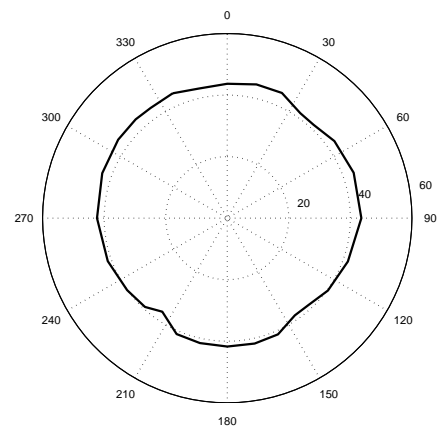


(f) パターン 4 の抽出結果

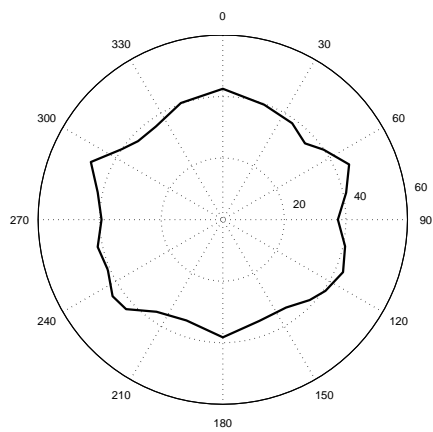
図 5.16: 方向ごとの放射指向特性の違いによるスピーカの音圧分布 (中心周波数:250 Hz)



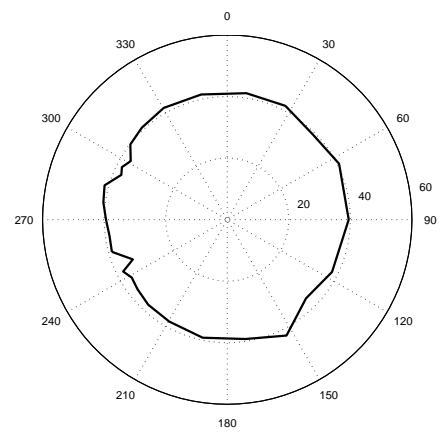
(a) 無響室での測定結果



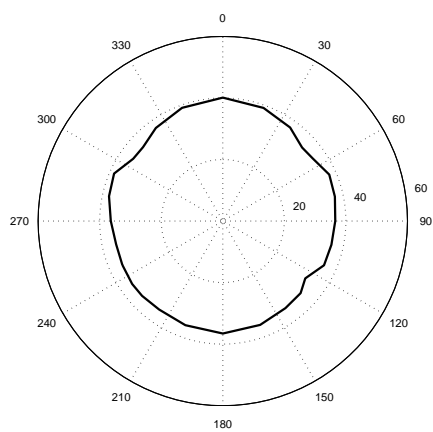
(d) パターン 2 の抽出結果



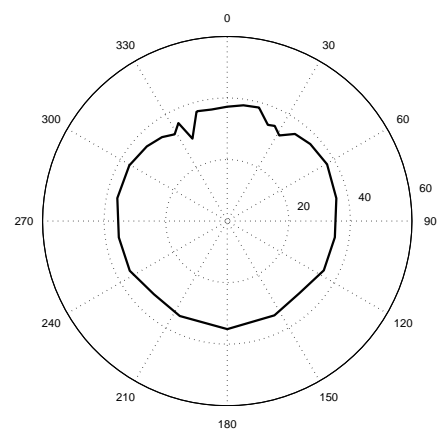
(b) 室内での測定結果 (パターン 1)



(e) パターン 3 の抽出結果



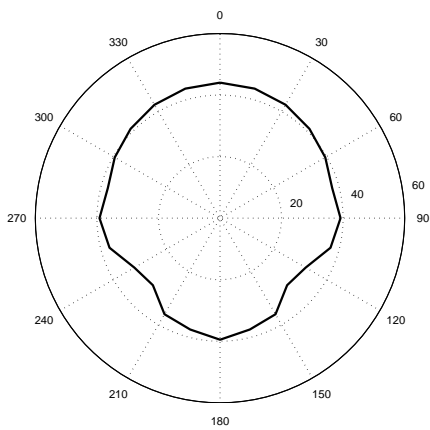
(c) パターン 1 の抽出結果



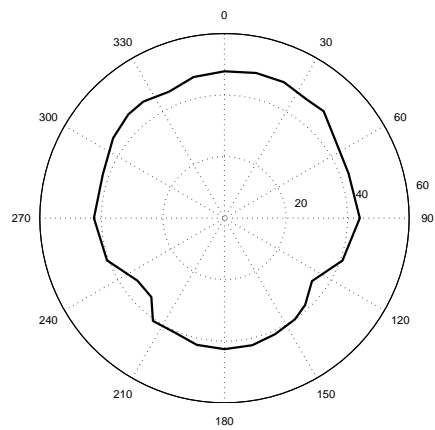
(f) パターン 4 の抽出結果

図 5.17: 方向ごとの放射指向特性の違いによるスピーカの音圧分布 (中心周波数:500 Hz)

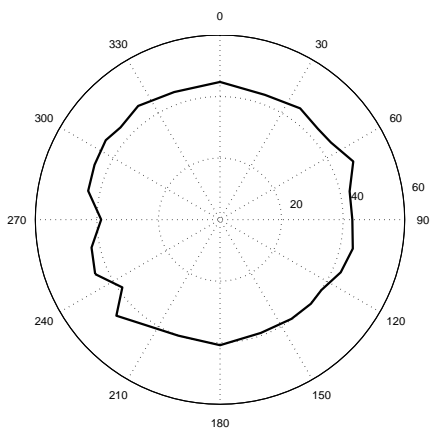




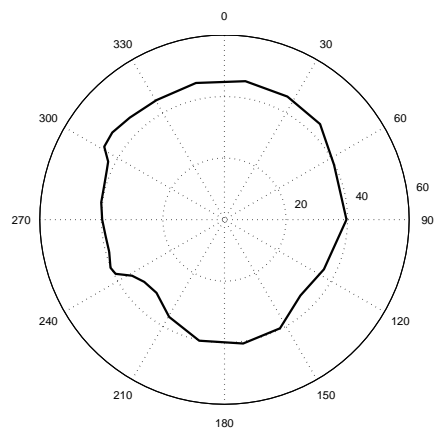
(a) 無響室での測定結果



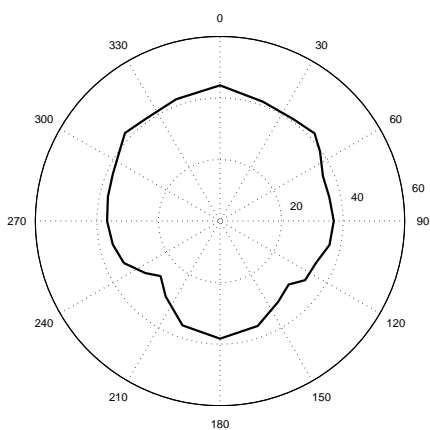
(d) パターン 2 の抽出結果



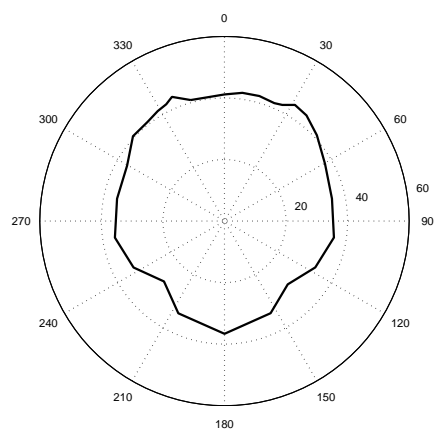
(b) 室内での測定結果 (パターン 1)



(e) パターン 3 の抽出結果

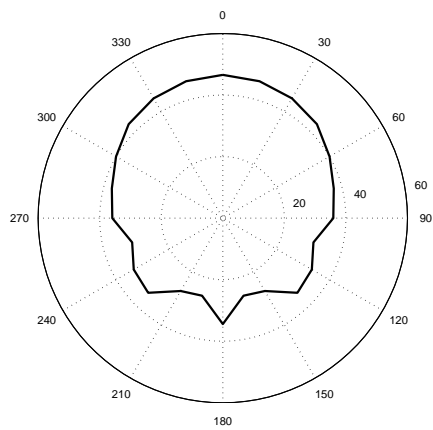


(c) パターン 1 の抽出結果

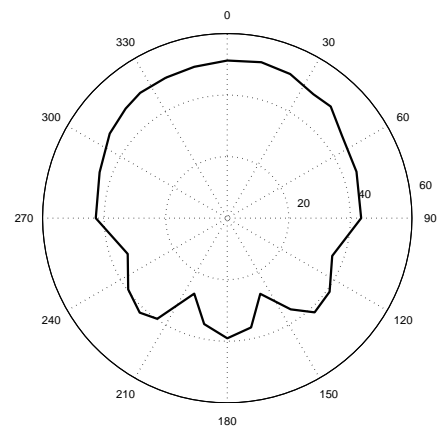


(f) パターン 4 の抽出結果

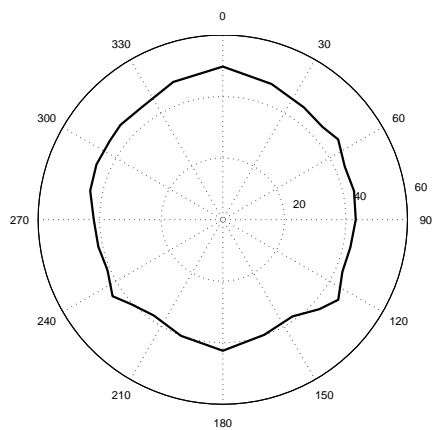
図 5.18: 方向ごとの放射指向特性の違いによるスピーカの音圧分布 (中心周波数: 1 kHz)



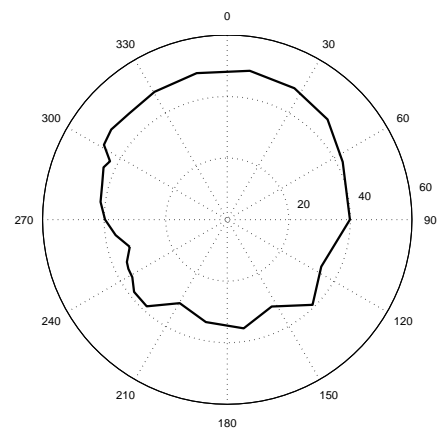
(a) 無響室での測定結果



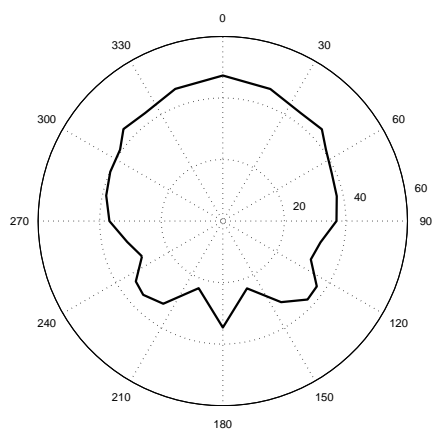
(d) パターン 2 の抽出結果



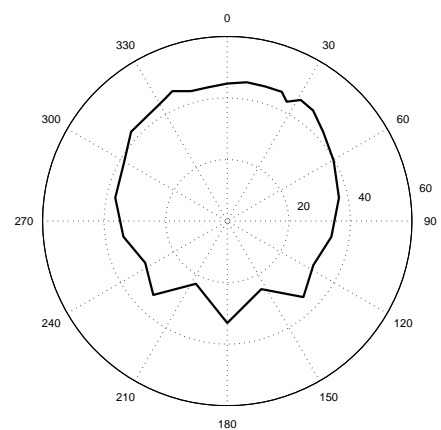
(b) 室内での測定結果 (パターン 1)



(e) パターン 3 の抽出結果

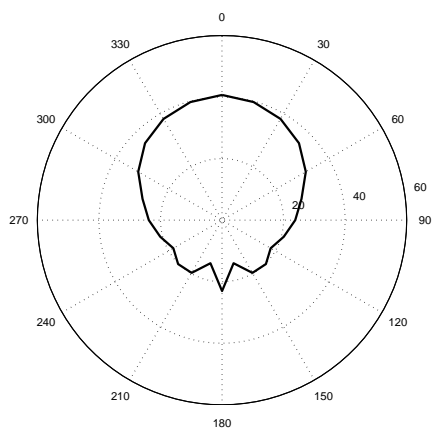


(c) パターン 1 の抽出結果

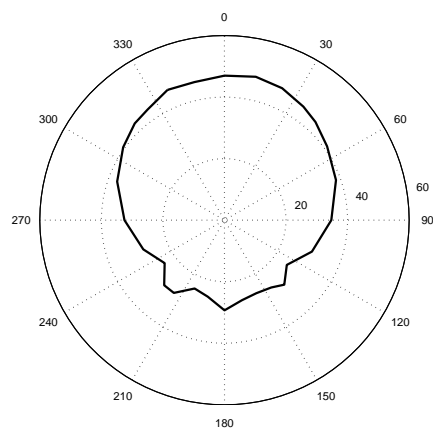


(f) パターン 4 の抽出結果

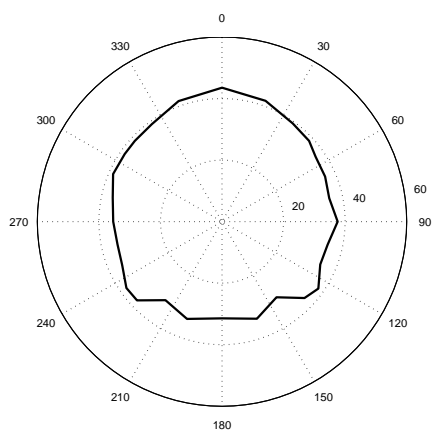
図 5.19: 方向ごとの放射指向特性の違いによるスピーカの音圧分布 (中心周波数: 2k Hz)



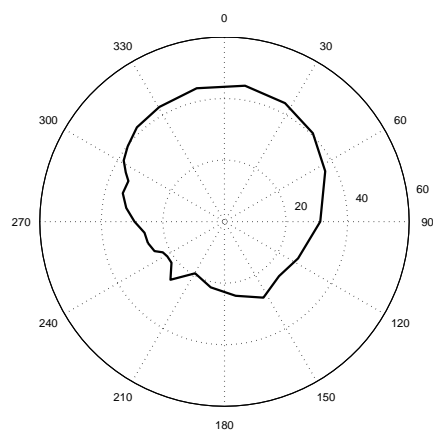
(a) 無響室での測定結果



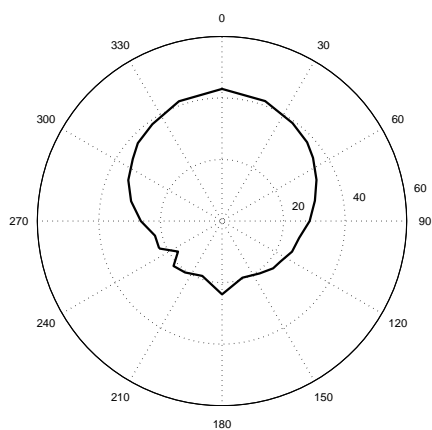
(d) パターン 2 の抽出結果



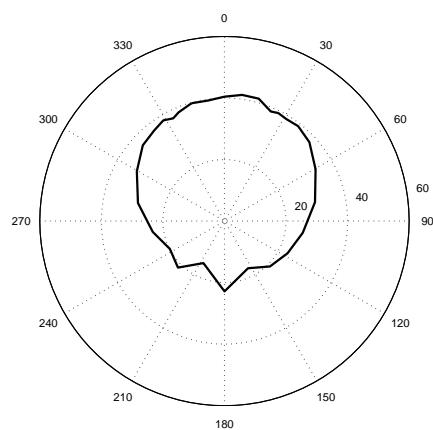
(b) 室内での測定結果 (パターン 1)



(e) パターン 3 の抽出結果

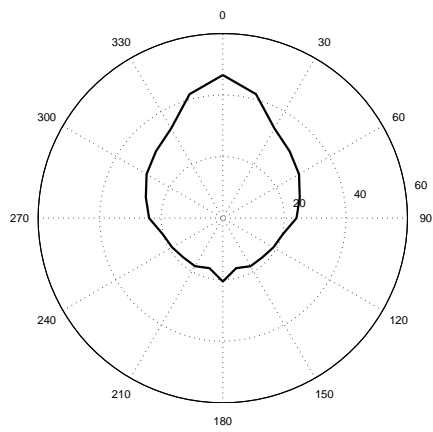


(c) パターン 1 の抽出結果

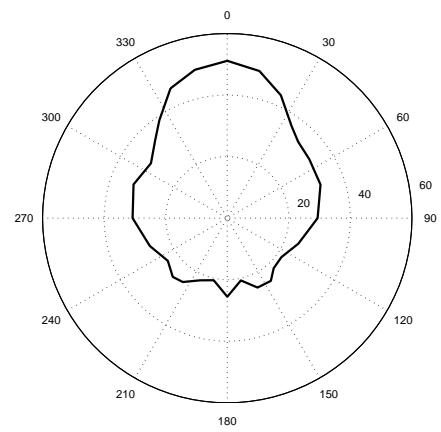


(f) パターン 4 の抽出結果

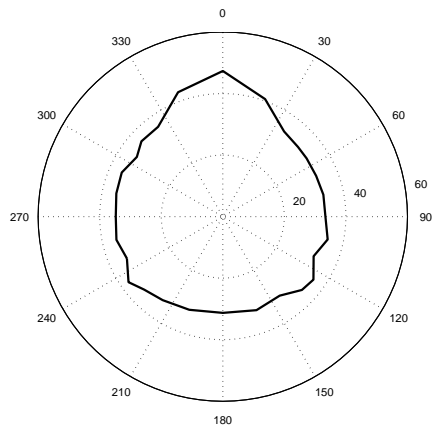
図 5.20: 方向ごとの放射指向特性の違いによるスピーカの音圧分布 (中心周波数: 4 kHz)



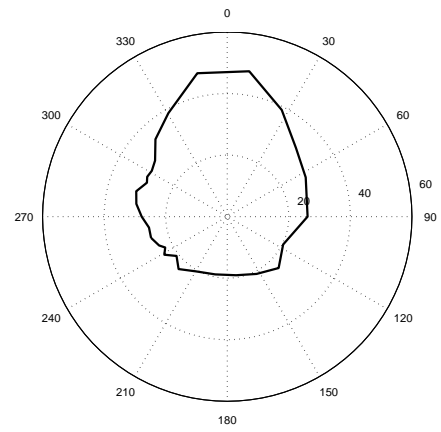
(a) 無響室での測定結果



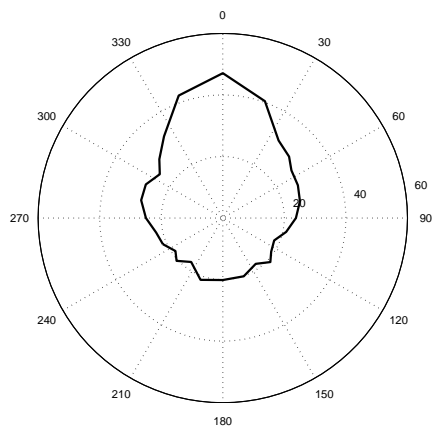
(d) パターン 2 の抽出結果



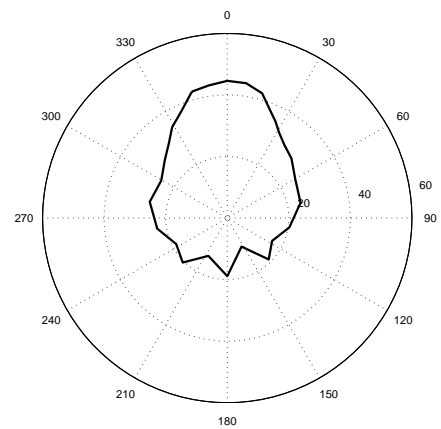
(b) 室内での測定結果 (パターン 1)



(e) パターン 3 の抽出結果

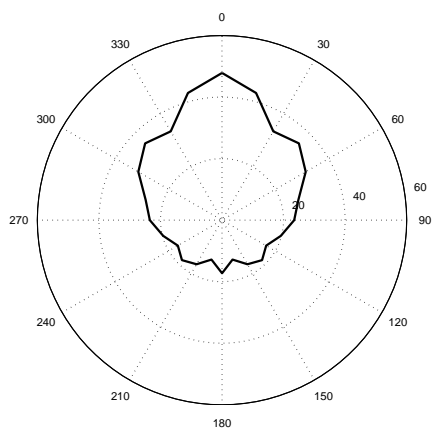


(c) パターン 1 の抽出結果

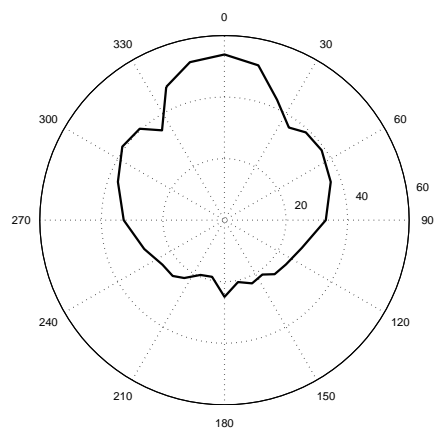


(f) パターン 4 の抽出結果

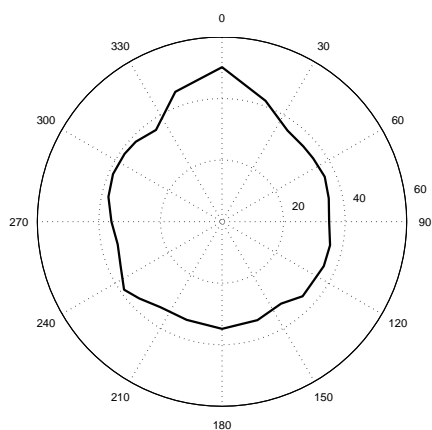
図 5.21: 方向ごとの放射指向特性の違いによるスピーカの音圧分布 (中心周波数: 8 kHz)



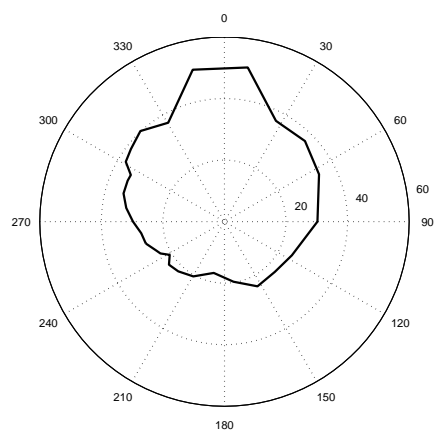
(a) 無響室での測定結果



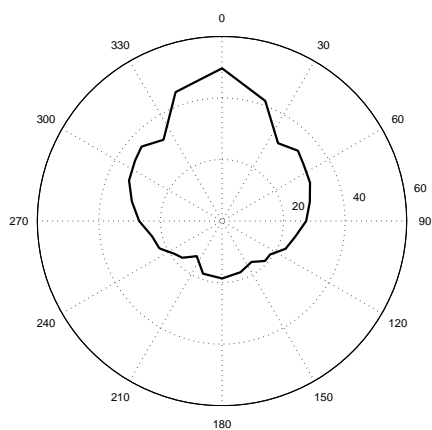
(d) パターン 2 の抽出結果



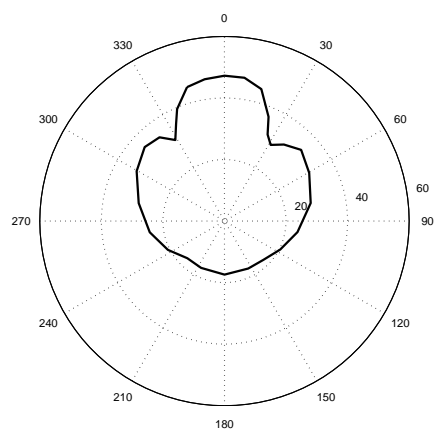
(b) 室内での測定結果 (パターン 1)



(e) パターン 3 の抽出結果

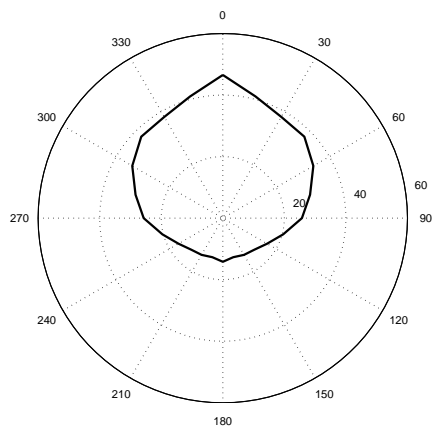


(c) パターン 1 の抽出結果

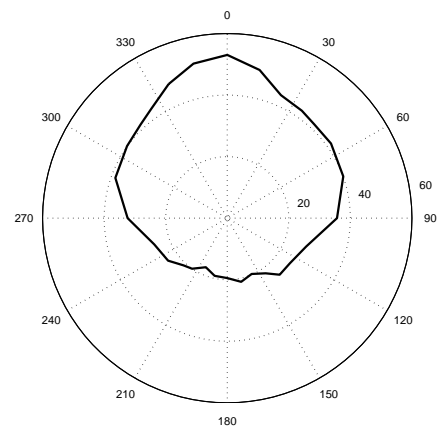


(f) パターン 4 の抽出結果

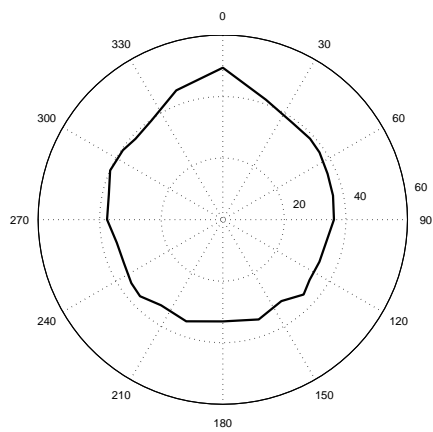
図 5.22: 方向ごとの放射指向特性の違いによるスピーカの音圧分布 (中心周波数:10 kHz)



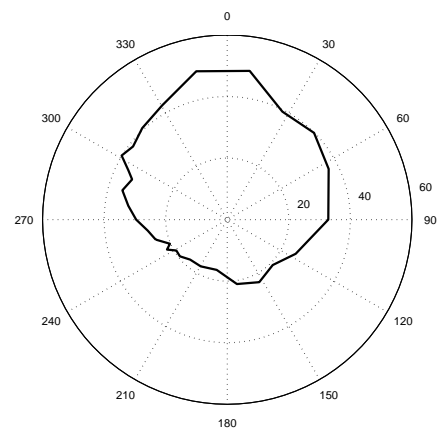
(a) 無響室での測定結果



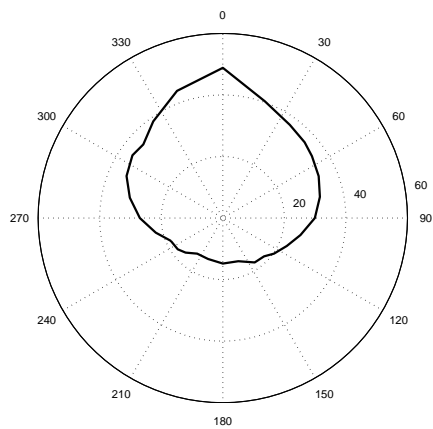
(d) パターン 2 の抽出結果



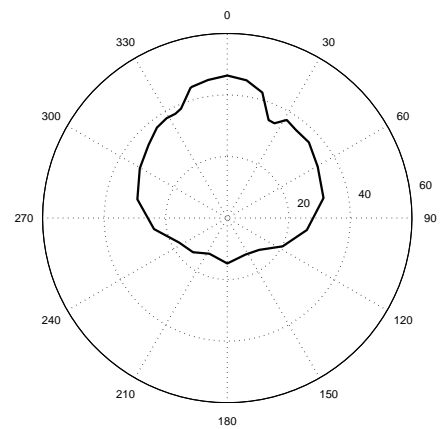
(b) 室内での測定結果 (パターン 1)



(e) パターン 3 の抽出結果

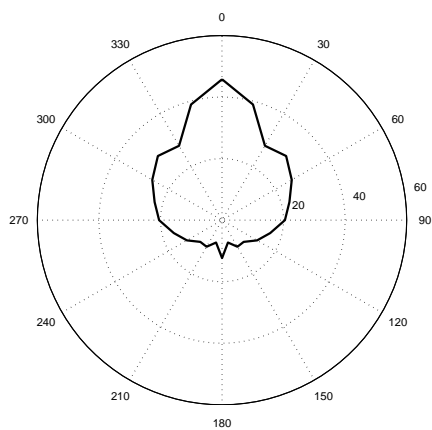


(c) パターン 1 の抽出結果

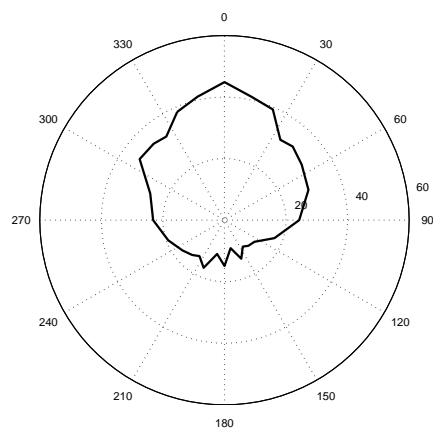


(f) パターン 4 の抽出結果

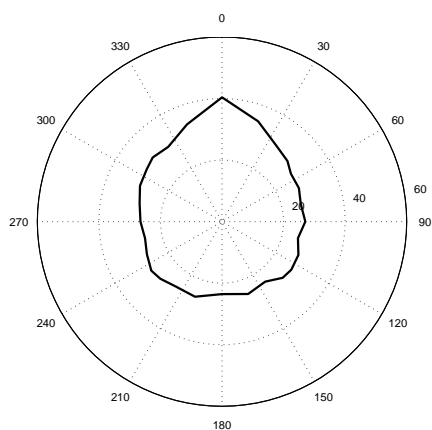
図 5.23: 方向ごとの放射指向特性の違いによるスピーカの音圧分布 (中心周波数 : 12.5 kHz)



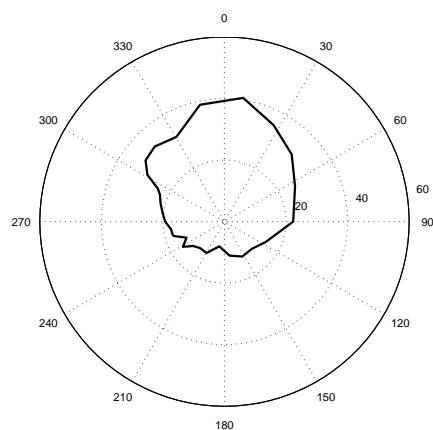
(a) 無響室での測定結果



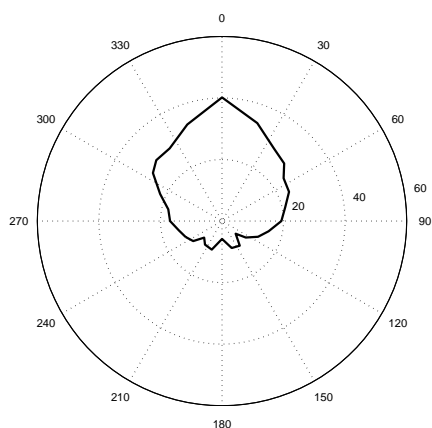
(d) パターン 2 の抽出結果



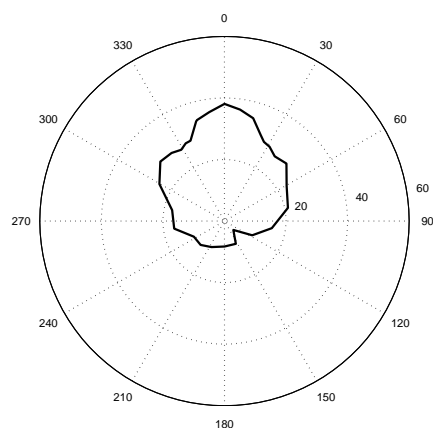
(b) 室内での測定結果 (パターン 1)



(e) パターン 3 の抽出結果

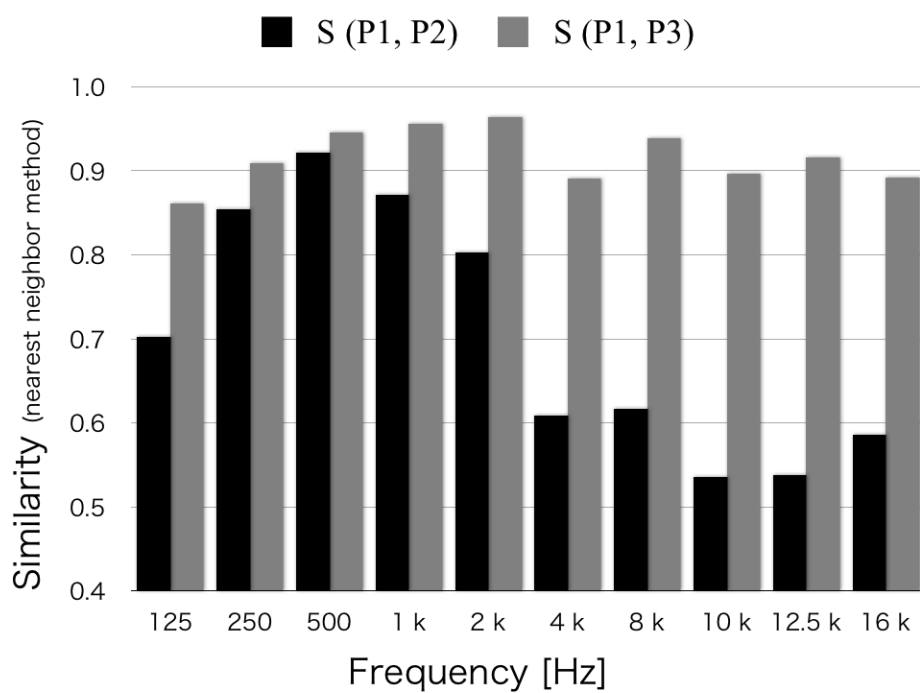


(c) パターン 1 の抽出結果

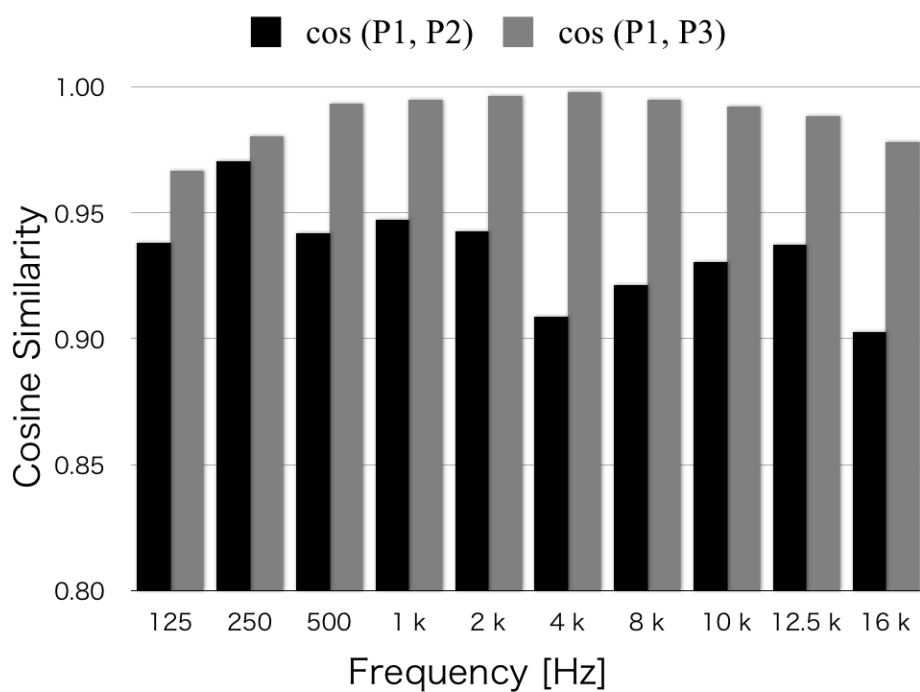


(f) パターン 4 の抽出結果

図 5.24: 方向ごとの放射指向特性の違いによるスピーカの音圧分布 (中心周波数:16 kHz)



(a) 最短距離法に基づく類似度

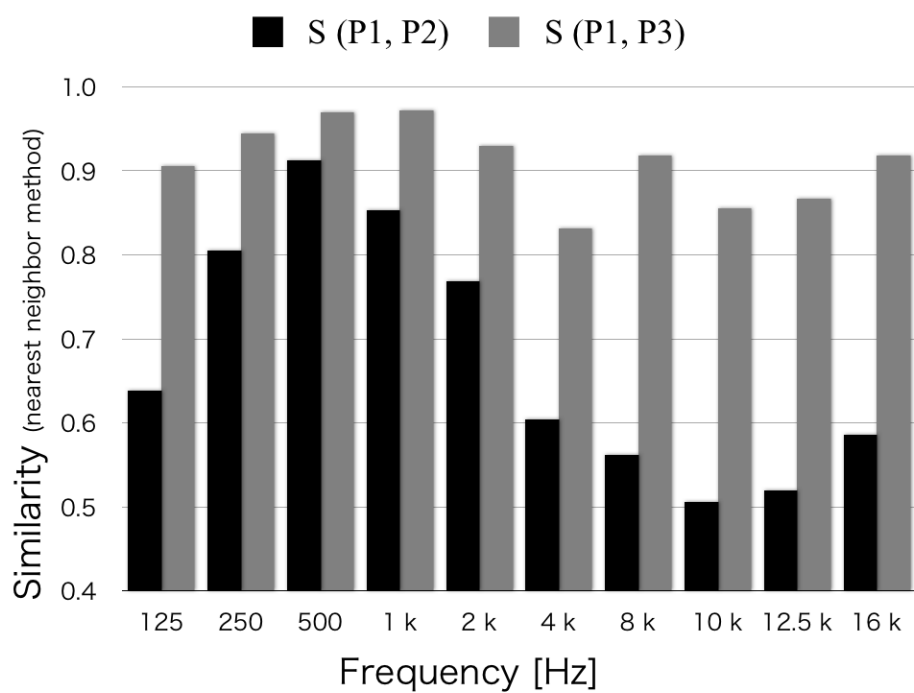


(b) 余弦尺度

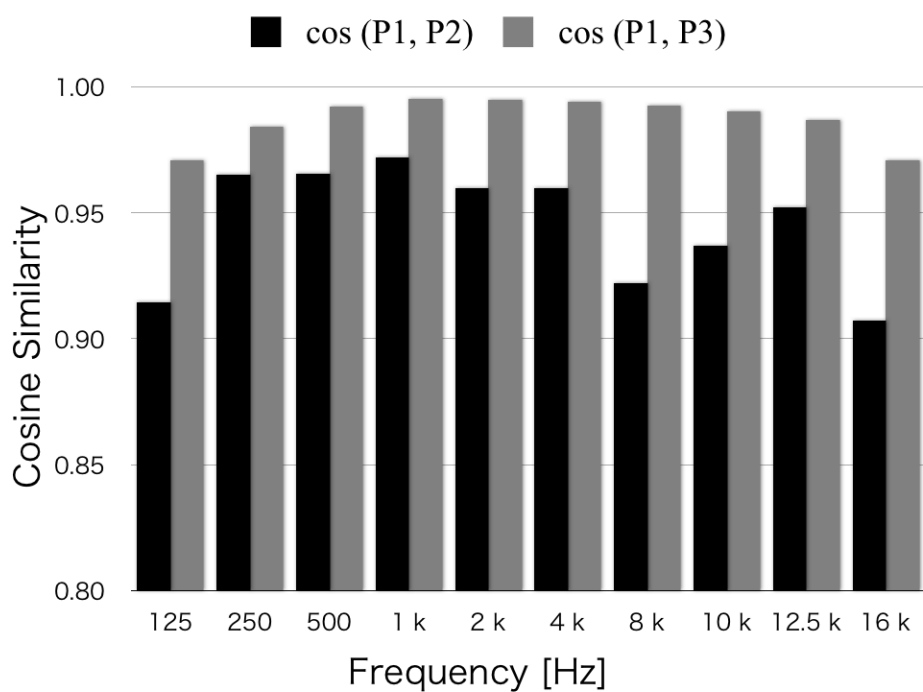
図 5.25: 各バンドごとの音圧分布の類似度 (パターン 1)

$P_1$ : 無響室の測定結果の音圧分布のパターンベクトル,  $P_2$ : 室内の測定結果の音圧分布のパターンベクトル,  $P_3$ : 推定した放射指向特性の音圧分布のパターンベクトル





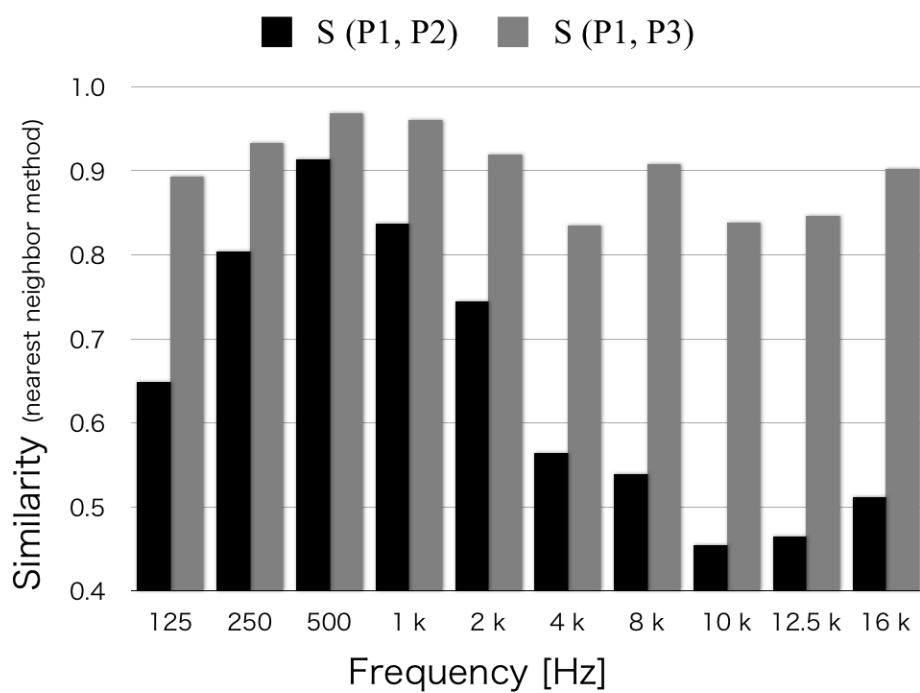
(a) 最短距離法に基づく類似度



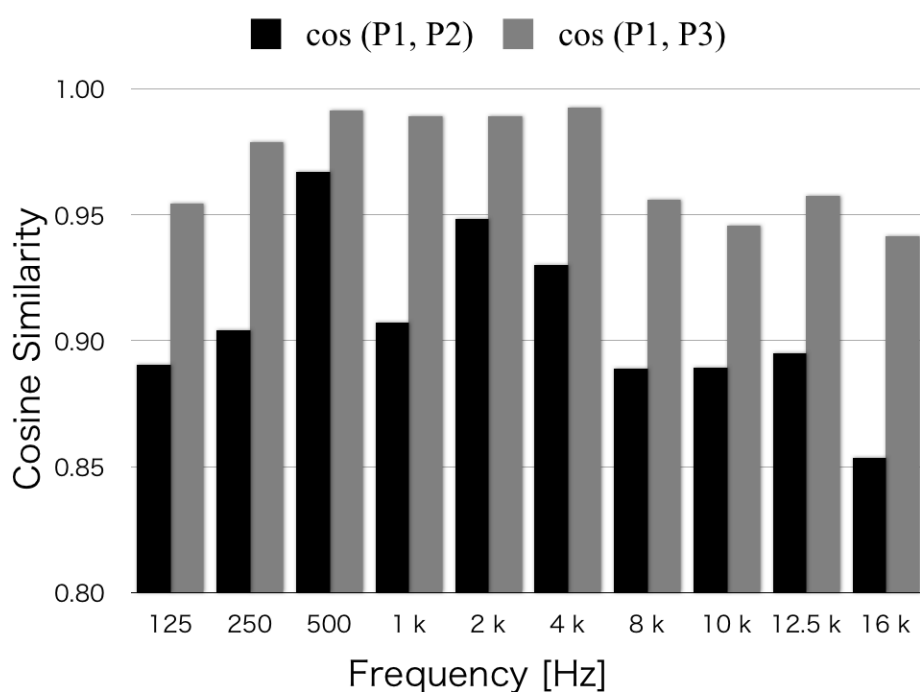
(b) 余弦尺度

図 5.26: 各バンドごとの音圧分布の類似度 (パターン 2)

$P_1$ : 無響室の測定結果の音圧分布のパターンベクトル,  $P_2$ : 室内の測定結果の音圧分布のパターンベクトル,  $P_3$ : 推定した放射指向特性の音圧分布のパターンベクトル



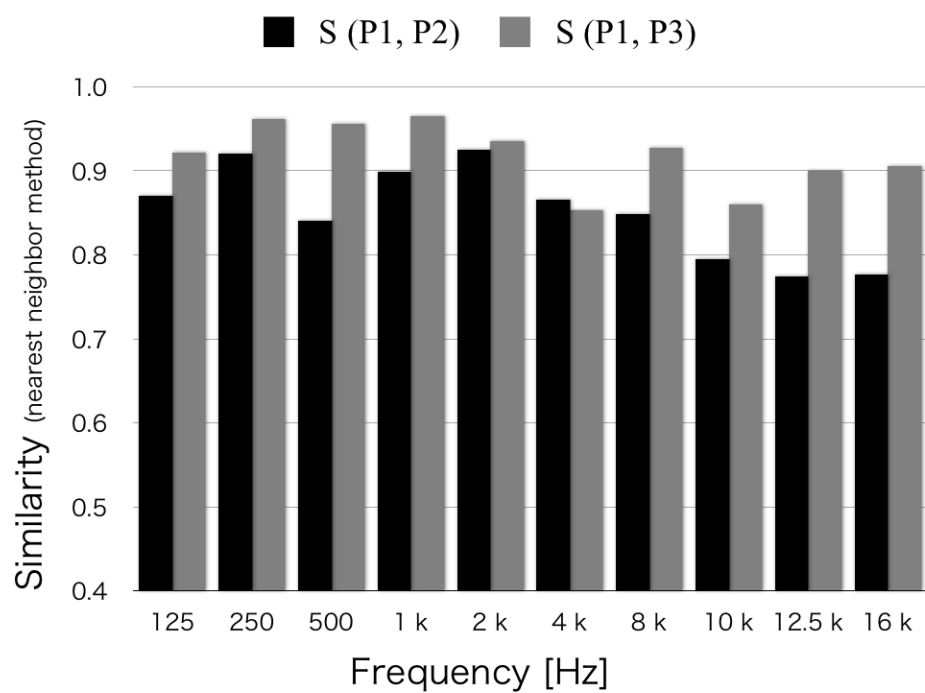
(a) 最短距離法に基づく類似度



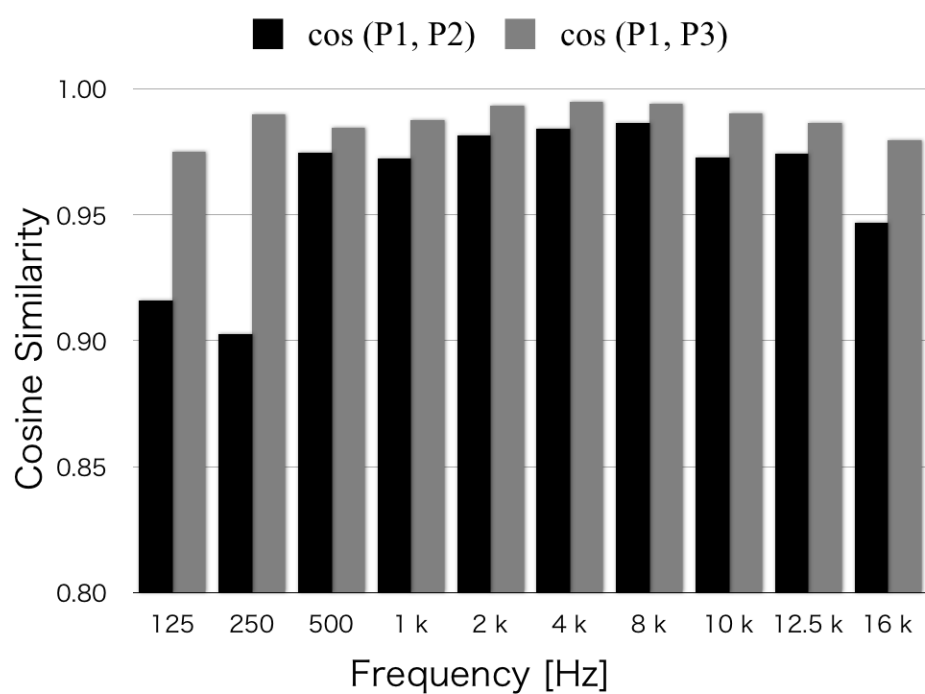
(b) 余弦尺度

図 5.27: 各バンドごとの音圧分布の類似度 (パターン 3)

$P_1$ : 無響室の測定結果の音圧分布のパターンベクトル,  $P_2$ : 室内の測定結果の音圧分布のパターンベクトル,  $P_3$ : 推定した放射指向特性の音圧分布のパターンベクトル



(a) 最短距離法に基づく類似度



(b) 余弦尺度

図 5.28: 各バンドごとの音圧分布の類似度 (パターン 4)

$P_1$ : 無響室の測定結果の音圧分布のパターンベクトル,  $P_2$ : 室内の測定結果の音圧分布のパターンベクトル,  $P_3$ : 推定した放射指向特性の音圧分布のパターンベクトル

## 第6章 結論

本研究は、音空間全体を記録し、記録した音空間を構成する属性(音源位置情報、原音情報、放射指向特性、反射音、残響特性)を分解・抽出し、必要なものを自由に編集、提示を可能とする次世代音空間提示システムの構築を見すえた先験的研究として取り組んだものである。

序論では、実環境の音空間を構成する属性について述べ、現状の音空間提示システムの概要と音空間を構成する属性との関係について論じた。従来の音場再現が音空間全体を記録、提示するシステムであるのに対して、音空間を構成する属性の分離や属性ごとの提示が可能となれば、従来法では表現できなかった多様性に富んだ臨場感を表現できることを示し、本研究が最終的に目指す次世代音空間提示システムを考案した。その実現のために、(a) 音空間を記録する技術、(b) 記録した音空間から属性を分離、抽出する技術、(c) 属性を忠実にレンダリングする技術の3つが必要となることを述べ、本研究では(a)と(b)について取り組むこととした。その上で本研究で具体的に検討する事項として、室内音空間全体を收音するシステムの構築と、反射音を含んだ音情報から、音源の属性である音源位置情報、原音情報、放射指向特性の抽出を挙げた。その中でも、放射指向特性は音源の方向ごとの空間情報を提示するために重要である。しかし、これまで音情報の記録や提示に関する研究では、音源の放射指向は考慮されていないことが多い。いわば、放射指向特性の抽出は、本研究の最大の目標とすることができる。また、記録した反射音を含む音情報から音源位置情報、原音情報、放射指向特性を抽出するために解決すべき課題と論文全体の章構成について述べた。

第2章では、音空間全体を記録するためのシステムの構築に関して述べた。具体的には、室内の床を除く壁面4面と天井に、マイクロホンを格子状に配置した包囲型マイクロホンアレイシステムを実装した。マイクロホンアレイではチャンネル全体での同期が必須である。本システムでは157 chと非常に多いチャンネル数であるため、クロックジェネレータとソフトウェアによる同期システムを用いて複数のPCを用いた同時録音を可能とした。TSP信号を用いて同期の確認を行い、157 ch、サンプリング周波数48 kHz、16 bit リニア PCMの同時録音が可能であることを示した。また、第4章と第5章の検討に必要な、包囲型マイクロホンアレイを用いた室内インパルスの測定方法についても言及した。

第3章では、反射音に頑健な音源位置推定に関する検討を行った。原音情報、そして放射指向特性を抽出するためには、最初に音源位置を高精度に求める必要がある。本研究では反射残響の存在する室内を前提としている。反射音は直接音と相関が高い波であるため、音源位置推定精度が劣化することが問題となる。そこで、第3章では、反射音に頑健な音源位置推定に関する検討を行った。従来法であるビームフォーミングに基づく方法の場合、マイクロホン間隔によって空間的な折り返し歪みが生じるという問題があるため、本研究では、マイクロホン間隔に依存せず、音源位置を高精度に推定できる MUSIC に着目した。MUSIC においても直接音と相関の高い反射音が存在する環境においては推定精度が劣化するといった問題があり、従来法では空間平均化法によって反射音を抑圧する方法が提案されていた。しかし、この方法は3次元位置推定にそのまま適用することができなかった。そこで、空間平均化法を3次元位置推定に導入するために、仮想チャネルを探索点ごとに生成し直す、即ちステアリングベクトルを探索点ごとに計算し直す方法を提案した。さらに、観測信号そのものを直接平均化し、空間平均化法よりも反射音の抑圧効果の高い RAP-MUSIC を考案した。実環境における音源位置推定実験の結果、提案法である RAP-MUSIC の方が、反射音の存在する環境において従来法よりも精度よく音源位置を推定できることを確認した。

第4章では、観測信号から反射音を取り除き、原音を高精度に抽出する残響除去に関する検討を行った。高精度な残響除去を行う場合、ビームフォーミングによる原音強調よりも、室内インパルスの逆フィルタに基づく方法の方が残響を完全に除去できることから、本検討においても逆フィルタに基づく検討を行うこととした。まず最初に、多チャネル逆フィルタの基本原理である MINT について述べ、MINT 型逆フィルタを形成するための方法としてブラインドシステム同定とブラインド残響除去があり、それぞれに関する先行研究と特徴について述べた。室内インパルスは応答長が非常に大きいことや、応答長が不明であることから、それらの点を克服しているブラインド残響除去に関する検討を行うこととした。高精度に残響を除去できるブラインド残響除去法として LIME アルゴリズムを挙げ、この方法が残響時間が長い場合やサンプリング周波数が高い場合には、残響除去精度が劣化するという問題を指摘した。残響除去の検討の多くは後段に音声認識を想定しているため、16 kHz までの低サンプリング周波数が使用されているが、本研究では、物音や楽器音といったさらに広帯域の信号の残響除去も想定しているため、高サンプリング周波数においても良好な残響除去を可能とする方法の開発を行った。従来の LIME アルゴリズムは、観測信号の周波数軸方向のスパース性によって観測信号の相関行列のランクが低下し、原音の AR 係数を正しく推定できなくなることが問題であることを示した。この問題に対して、観測信号をあらかじめ白色化してから LIME アルゴリズムを適用し、残響除去された信号を白色化したフィルタで逆畳み込みすることによって、入力信号のスパース性問題を解決

し、高精度な残響除去が可能である新しい信号処理法、White-LIME を考案した。従来法と観測信号の相関行列のランクを比べた結果、考案した White-LIME ではランクの低下が少なく、高精度に残響除去を行うことができることを確認した。シミュレーションによる音声信号と音楽信号の残響除去の結果から、観測信号の周波数領域におけるスパース性が顕著となる高サンプリング時においても、提案法により精度よく残響除去ができることを確認した。以上、第 4 章の検討によって、高精度な残響除去による原音抽出が可能となった。

第 5 章では、本研究の最大の目的である音源の放射指向特性の抽出に関する検討を行った。放射指向特性を属性値として抽出するためには、放射指向特性をインパルス応答として抽出することが必要である。しかし、音源と各観測点のインパルス応答は放射指向特性と室内伝達特性が混合されたものとなっている。そこで本研究では、まず最初に第 4 章で考案した White-LIME によって抽出した原音情報を使って音源位置と観測点間のインパルス応答を抽出し、その応答を放射指向特性成分と室内伝達特性成分とに分離し、放射指向特性のみを抽出するという方針を立てた。まずは放射指向特性そのものの特性を把握するために、無響室におけるスピーカの放射指向特性の測定を行い、その傾向を確認した。次に、実環境での測定結果と比較を行った。放射指向特性のインパルス応答は全体のインパルスの初期応答部分であり、室内インパルス応答はその後部に位置すると仮定した。全体のインパルス応答はそれらの時間軸での足し合わせであることを確認した。そこで分離方針としては、抽出したインパルス応答のうち、直接音の到来時間から初期反射の到来時間までの応答を放射指向特性の伝達関数として分離することとした。第 4 章で考案した White-LIME で抽出した原音からインパルス応答を高精度に抽出できることを確認し、その結果を用いて放射指向特性の抽出を行った。この際、音源と各観測点までの距離が異なるため、放射指向特性として得るためには距離補正が必要であった。距離補正には、第 3 章で考案した RAP-MUSIC によって抽出した音源の位置情報と観測点の位置情報を用いた。抽出した放射指向特性の精度を確認するために、無響室で測定した放射指向特性、および、切り出しを行う前の室内インパルス応答から求めた放射指向特性との比較を行った。その結果、低周波をのぞき、良好な放射指向特性を抽出できることを確認した。この結果により、抽出したインパルス応答から初期応答を分離することによって、音源の放射指向特性の概形を抽出することに成功した。これら第 5 章の検討によって、観測信号に含まれる放射指向特性と室内残響特性から、音源の放射指向特性のみをインパルス応答という属性値として分離・抽出することが可能となった。

以上第 2 章から第 5 章までの検討により、音空間全体を記録することと、室内残響を含む音情報から (i) 音源位置情報、(ii) 原音情報、そして (iii) 放射指向特性という音源の持つ 3 つの属性を高精度に、そして属性値として抽出することができた。これに

よって、本研究の研究背景である次世代音空間提示システムにおける (c) 属性ごとを忠実にレンダリングする技術に必要な属性値のうち、音源の持つ属性値 (i) ~ (iii) の抽出が可能となった。したがって、抽出した (i) ~ (iii) については、別イベント間の属性値の入れ替えや編集が可能となった。よって、本研究の結果により、従来の音場再現では表現できなかった多様性に富んだ音空間の提示の可能性を広げることができたと言える。

また、(i) 音源位置情報、(ii) 原音情報、そして (iii) 放射指向特性の属性値としての抽出が可能となると、放射指向特性を含んだ聴覚ディスプレイ技術も即座に可能となる。位置センサを導入して音源の距離だけでなく、放射指向特性も位置情報によって切り替えることにより、実際の音源に近づく、周りを回るときと同様の音情報の提示が可能となる。よって本研究の成果は、従来よりも詳細な聴覚ディスプレイシステムの構築に貢献することができるであろう。

最後に、音空間からの完全な属性抽出に関するさらなる発展のための課題について述べる。本研究では、原音の抽出および放射指向特性抽出においては背景雑音の存在しない環境を想定し、シミュレーションによる検討を行ってきた。実環境への適用を考えた場合、観測信号に背景雑音  $b(n)$  が混入することは必須である。背景雑音が混入した場合、観測信号から直接 MINT 型逆フィルタを算出すると雑音の影響によって逆フィルタの精度が劣化するため、原音の抽出精度も劣化するといった問題がある。よって、背景雑音にロバストな方法 [104, 122] が検討課題である (図 6.1)。

実環境での適用での最大の課題としては、音源が複数ある Multi-Input Multiple-Output (MIMO) モデルでの検討であり、これはマイクロホンアレイシステムでの最大の検討課題である [123–125] (図 6.2)。

以上、本研究が次世代音空間提示という新しい試みに対する先駆けとなり、今後の音空間情報の記録および提示をより一層活性化させるものになったと期待して本論文の結びとする。

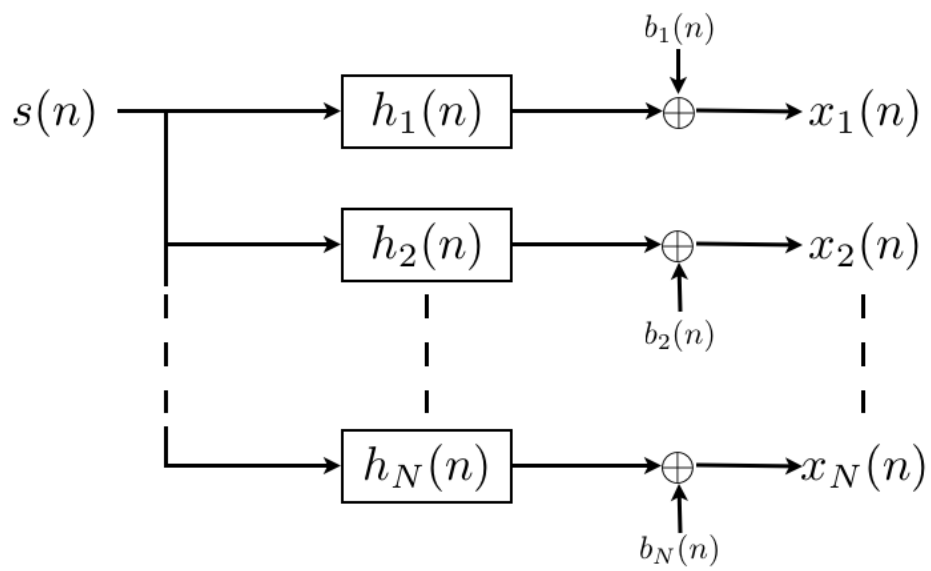


図 6.1: 背景雑音  $b_i(n)$  を考慮した SIMO モデル  $s(n)$  : 原音信号 ,  $h_i(n)$  : 室内伝達関数 ,  $x_i(n)$  : 観測信号



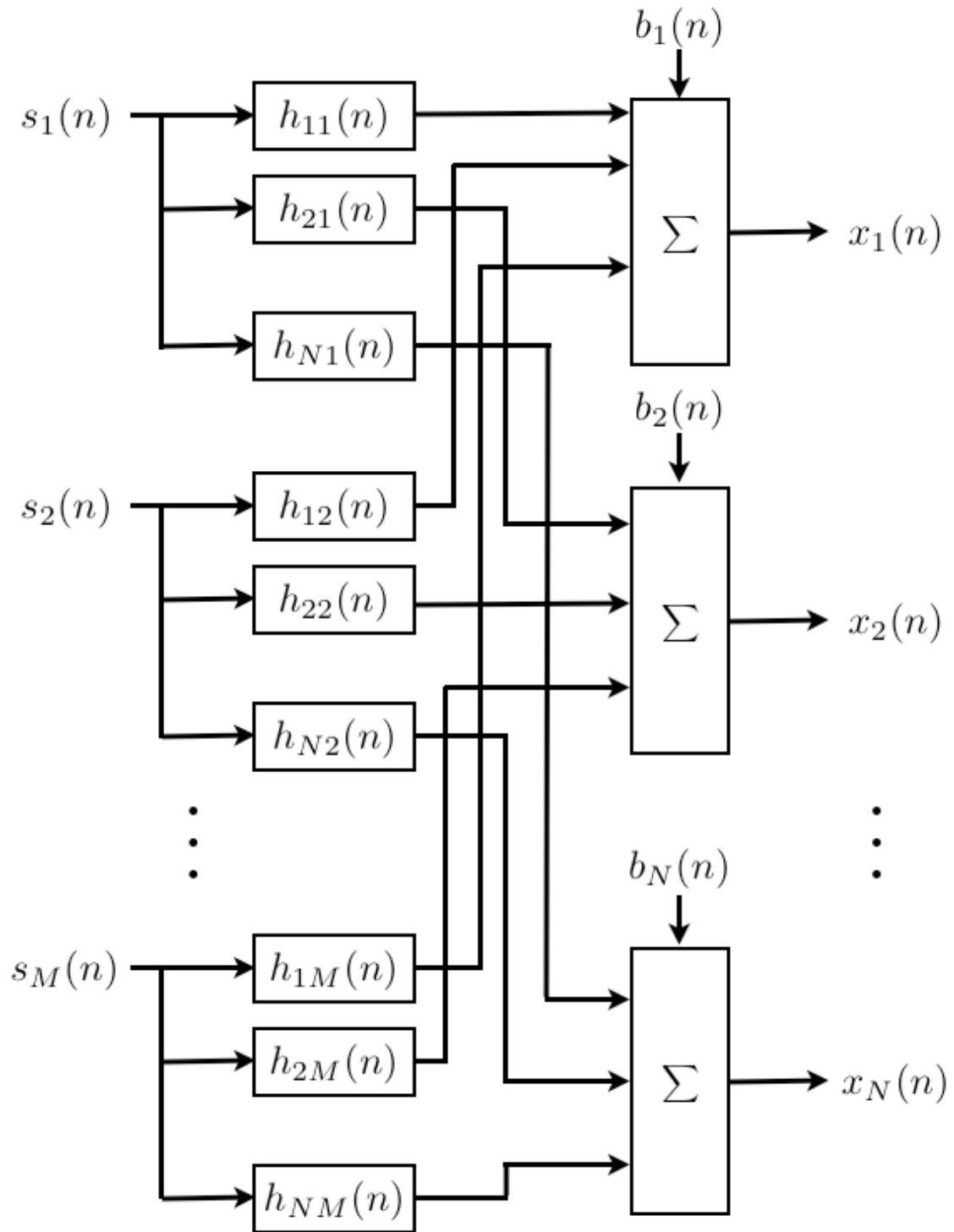
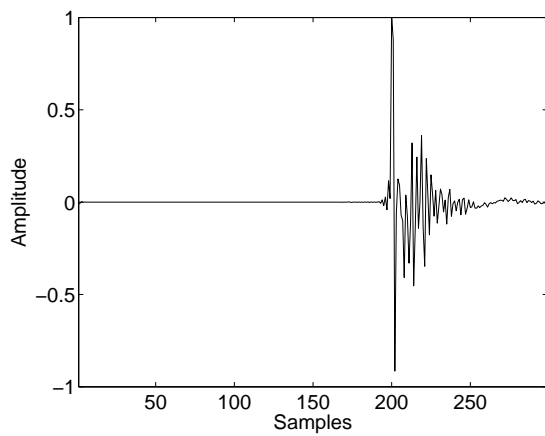


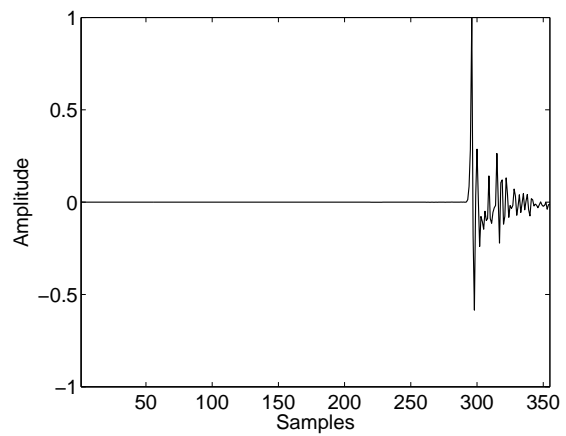
図 6.2: 背景雑音  $b_i(n)$  を考慮した SIMO モデル  $s_j(n)$ : 原音信号,  $h_{ij}(n)$ : 室内伝達関数,  $x_i(n)$ : 観測信号

## 付 録 A 放射指向特性の時間波形，周波数特性および位相特性

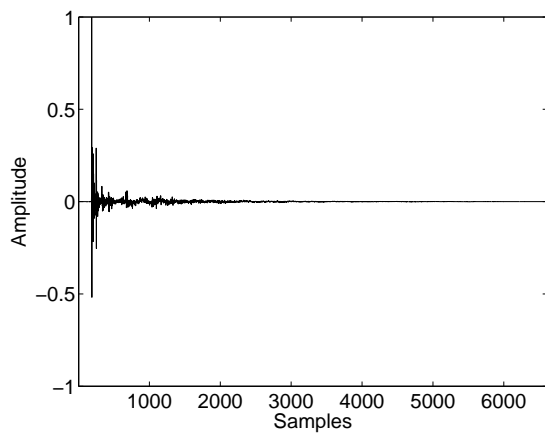
ここでは，第 5 章にて測定した 1 ウェイスピーカの放射指向特性の 0 °方向，90 °方向，180 °方向におけるインパルス応答，周波数特性および位相特性を示す．



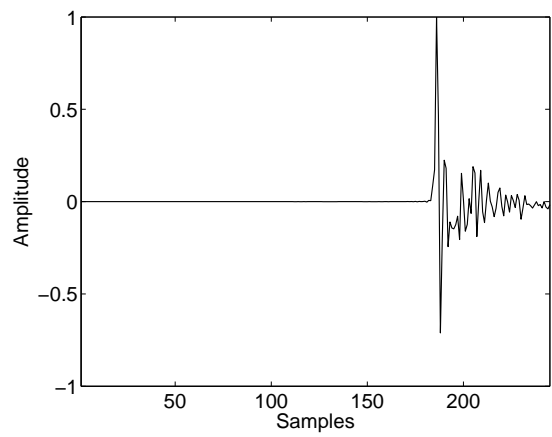
(a) 無響室での測定結果



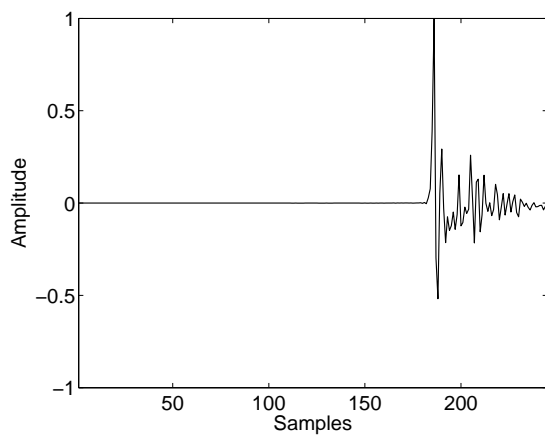
(d) パターン 2 の抽出結果



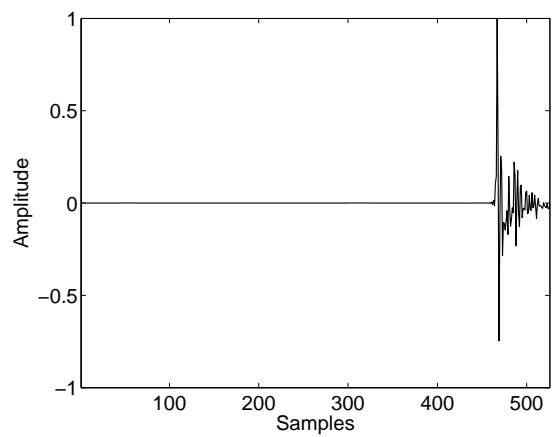
(b) 室内での測定結果 (パターン 1)



(e) パターン 3 の抽出結果

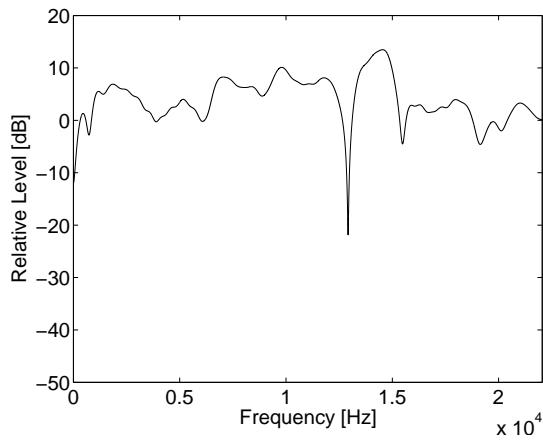


(c) パターン 1 の抽出結果

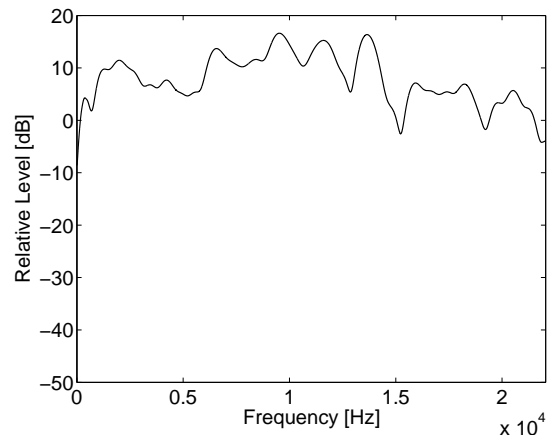


(f) パターン 4 の抽出結果

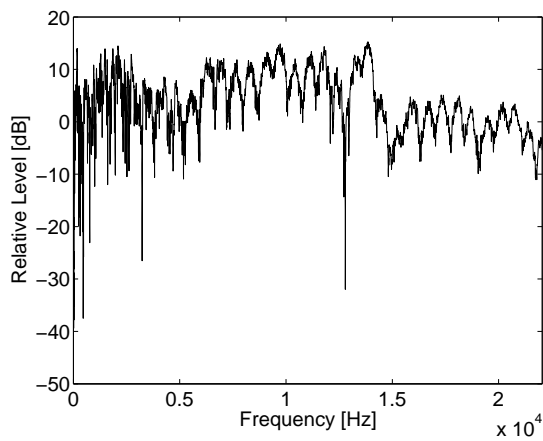
図 A.1: 0 °方向の放射指向特性：インパルス応答



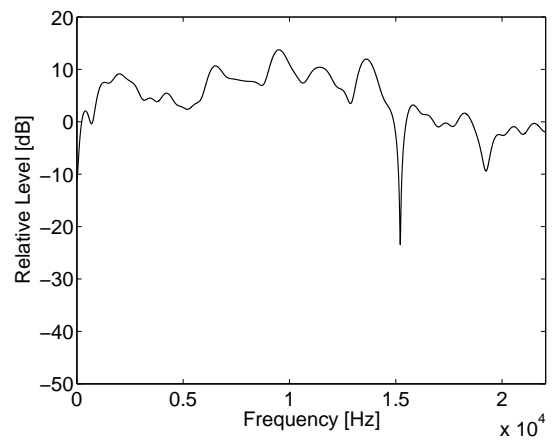
(a) 無響室での測定結果



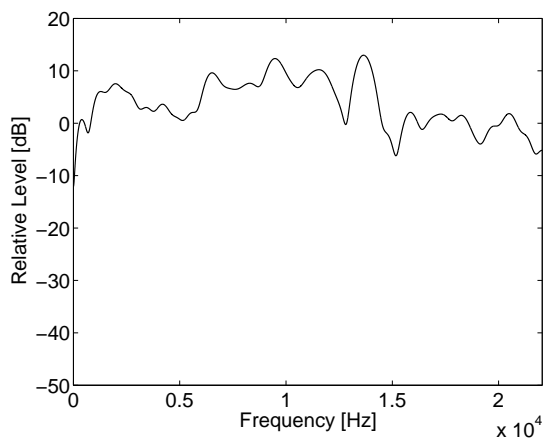
(d) パターン 2 の抽出結果



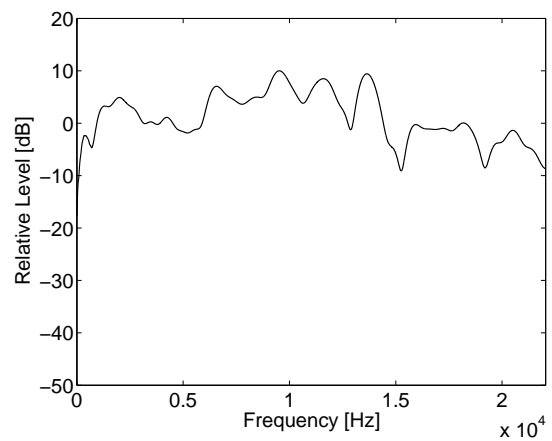
(b) 室内での測定結果 (パターン 1)



(e) パターン 3 の抽出結果

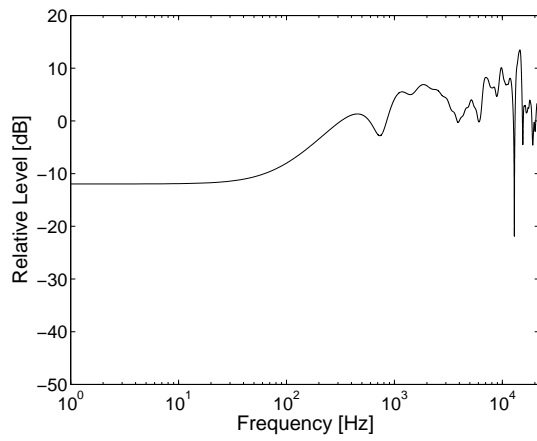


(c) パターン 1 の抽出結果

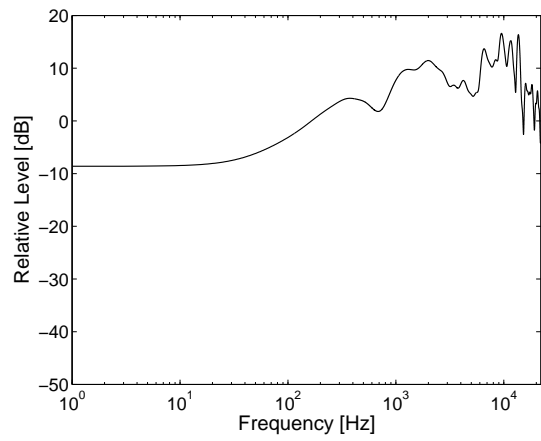


(f) パターン 4 の抽出結果

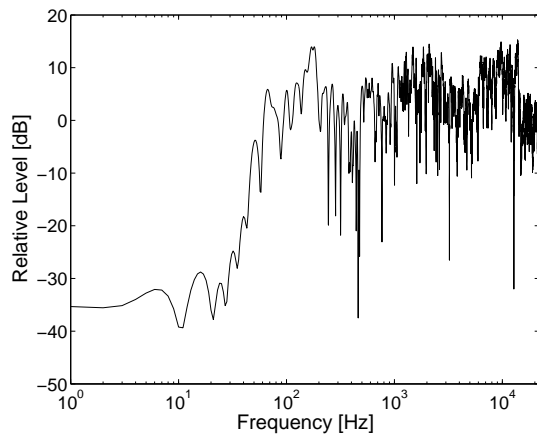
図 A.2: 0 °方向の放射指向特性：周波数応答



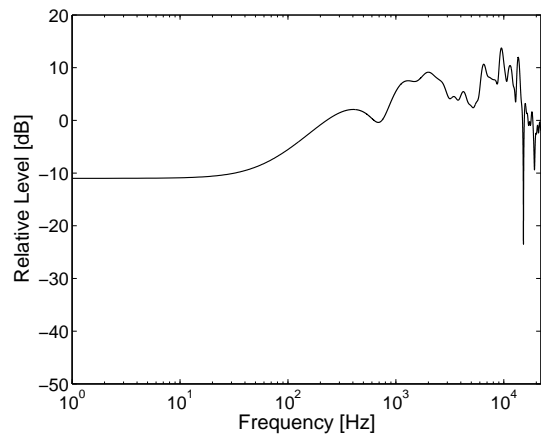
(a) 無響室での測定結果



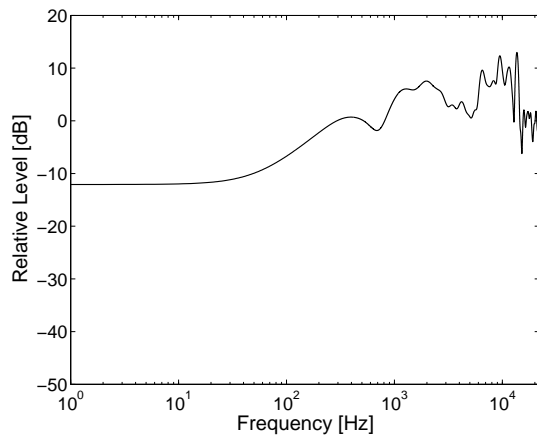
(d) パターン 2 の抽出結果



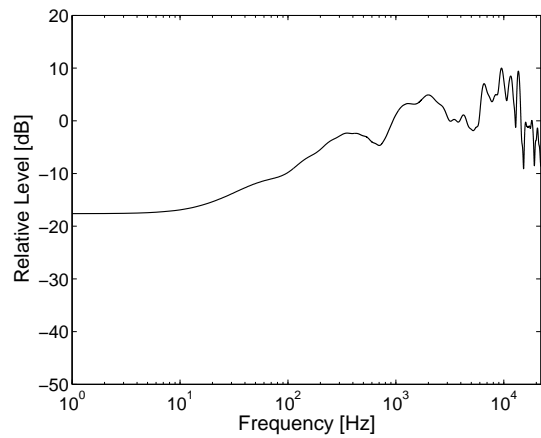
(b) 室内での測定結果 (パターン 1)



(e) パターン 3 の抽出結果

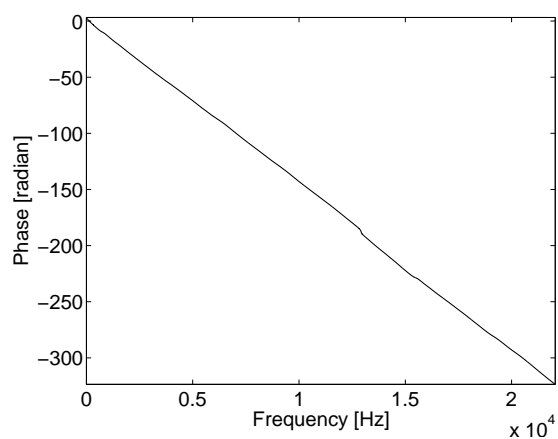


(c) パターン 1 の抽出結果

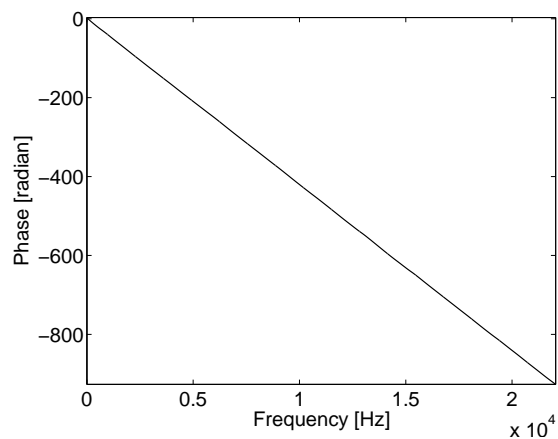


(f) パターン 4 の抽出結果

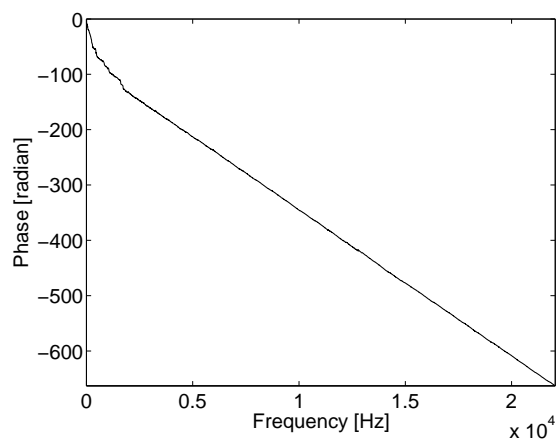
図 A.3: 0 °方向の放射指向特性：周波数応答 (対数スケール)



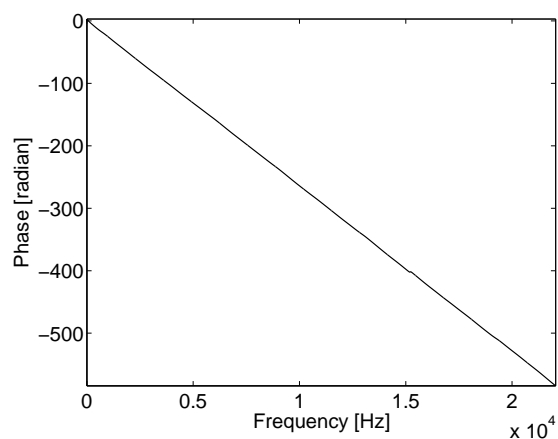
(a) 無響室での測定結果



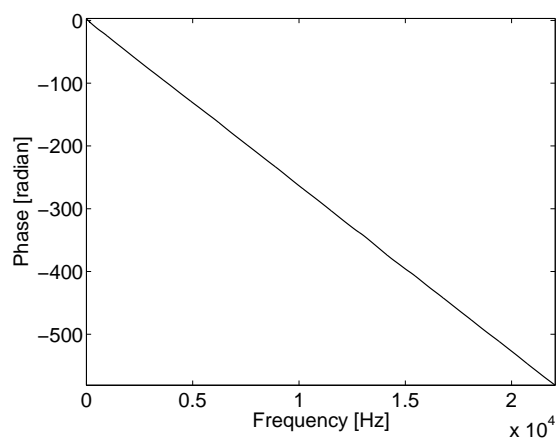
(d) パターン 2 の抽出結果



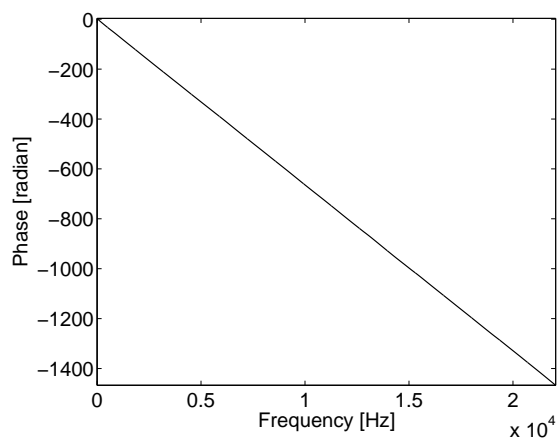
(b) 室内での測定結果 (パターン 1)



(e) パターン 3 の抽出結果

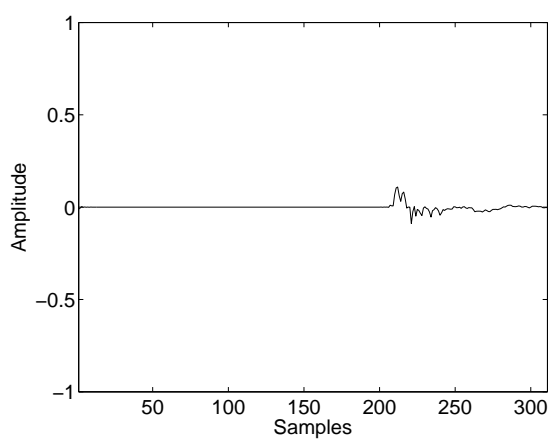


(c) パターン 1 の抽出結果

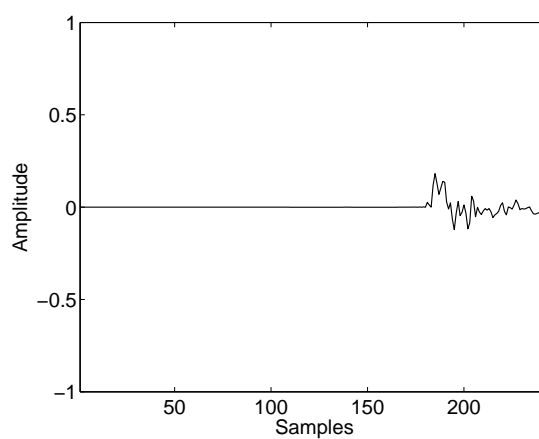


(f) パターン 4 の抽出結果

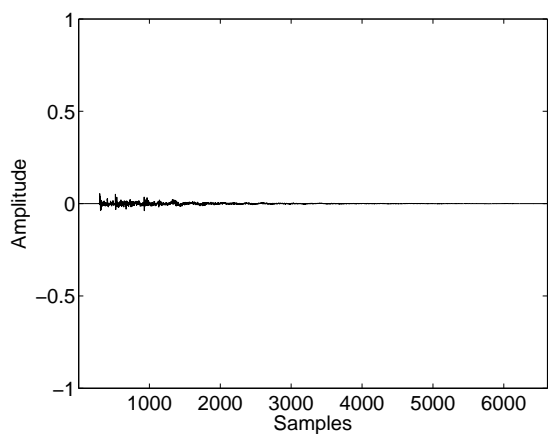
図 A.4: 0 °方向の放射指向特性：位相特性



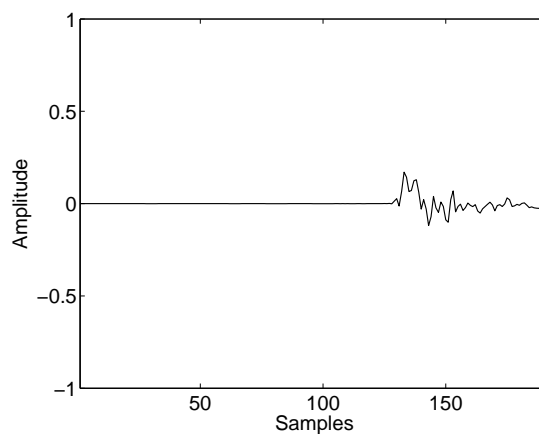
(a) 無響室での測定結果



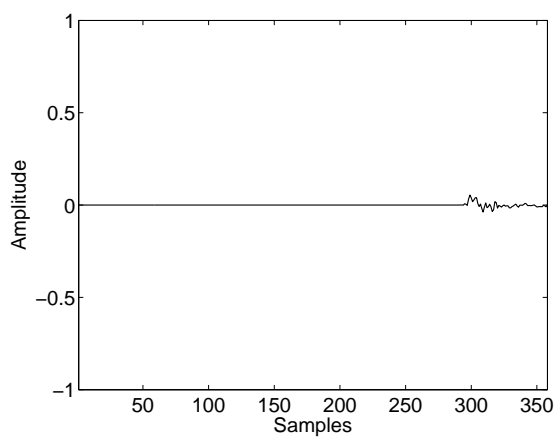
(d) パターン 2 の抽出結果



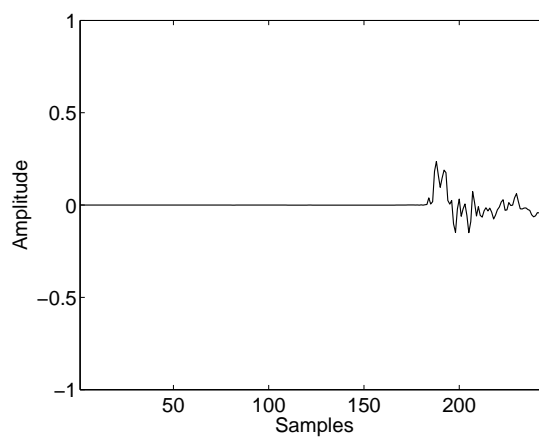
(b) 室内での測定結果 (パターン 1)



(e) パターン 3 の抽出結果

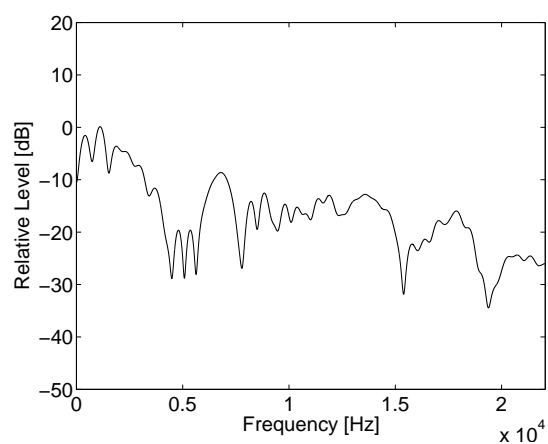


(c) パターン 1 の抽出結果

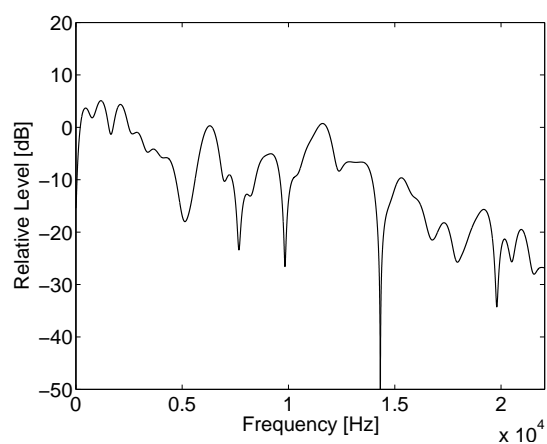


(f) パターン 4 の抽出結果

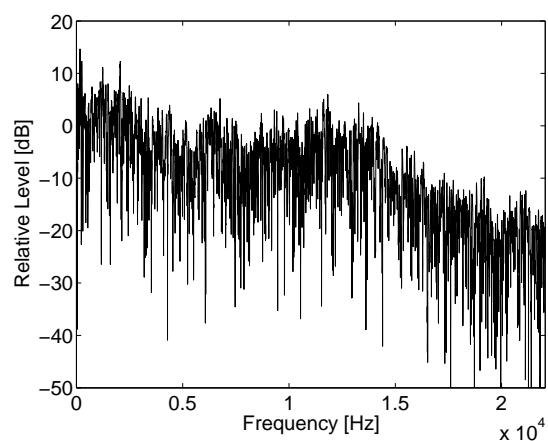
図 A.5: 90 °方向の放射指向特性：インパルス応答



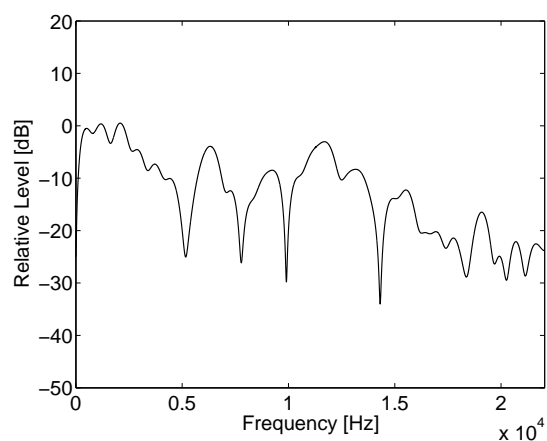
(a) 無響室での測定結果



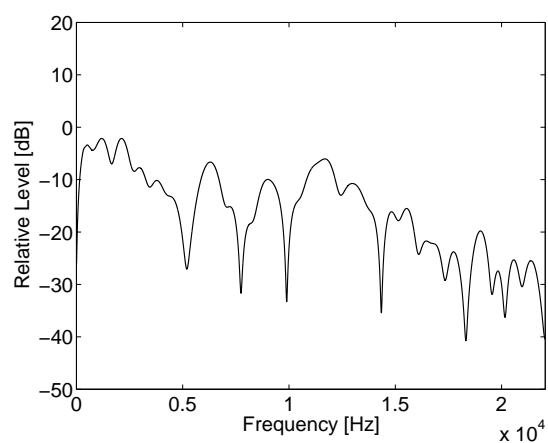
(d) パターン 2 の抽出結果



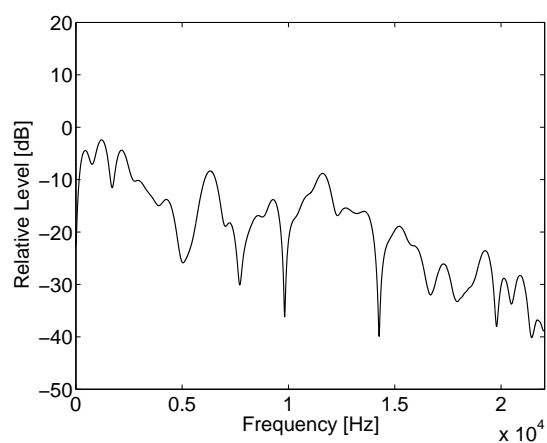
(b) 室内での測定結果 (パターン 1)



(e) パターン 3 の抽出結果



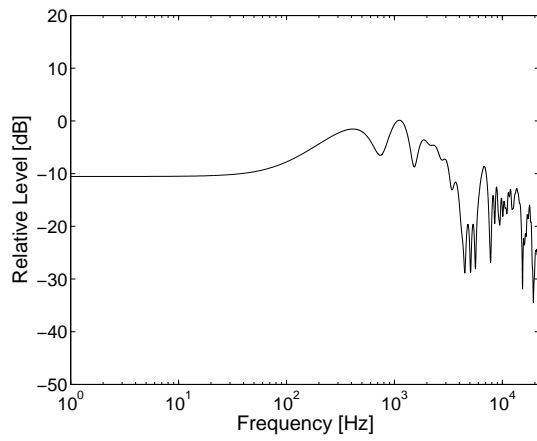
(c) パターン 1 の抽出結果



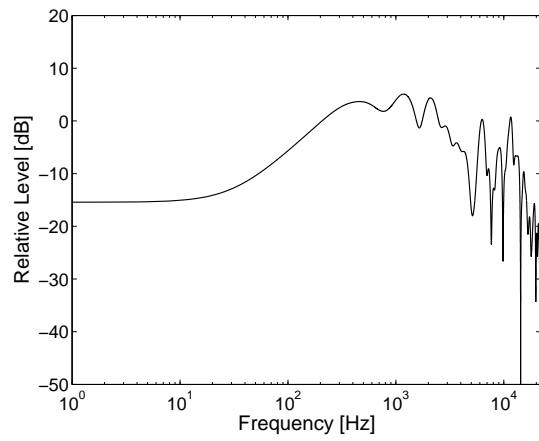
(f) パターン 4 の抽出結果

図 A.6: 90 °方向の放射指向特性：周波数応答

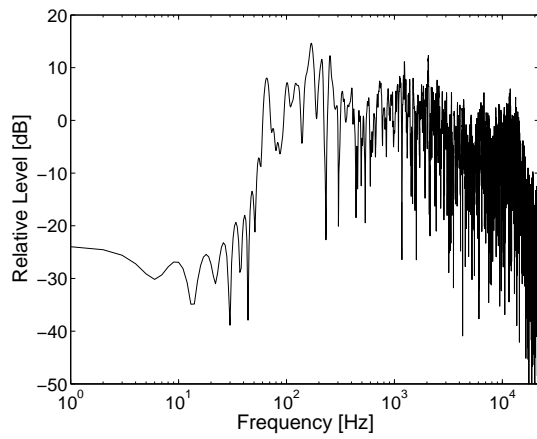




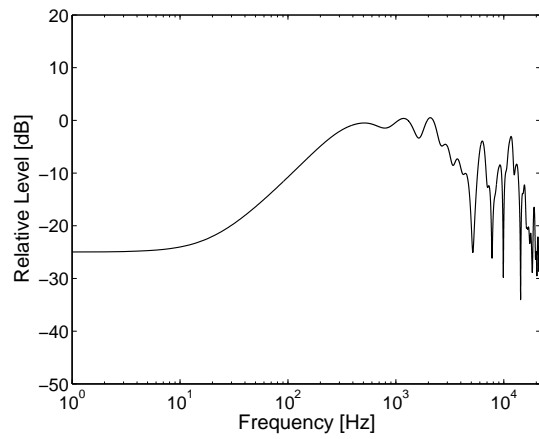
(a) 無響室での測定結果



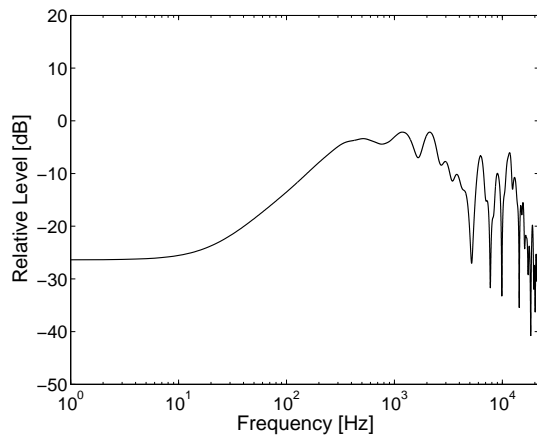
(d) パターン 2 の抽出結果



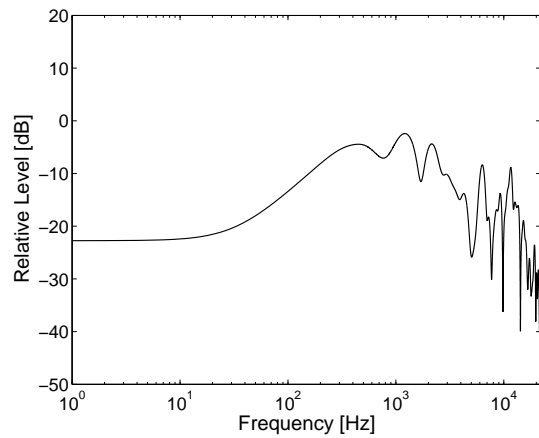
(b) 室内での測定結果 (パターン 1)



(e) パターン 3 の抽出結果

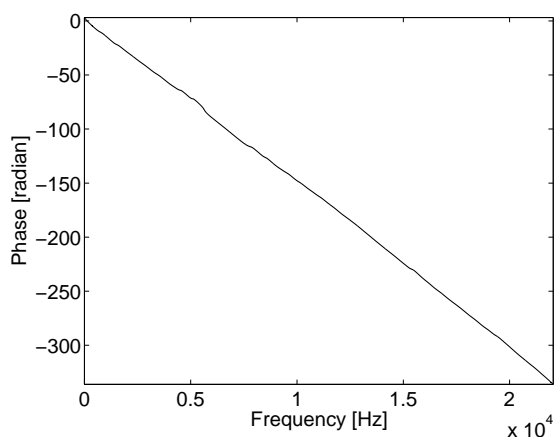


(c) パターン 1 の抽出結果

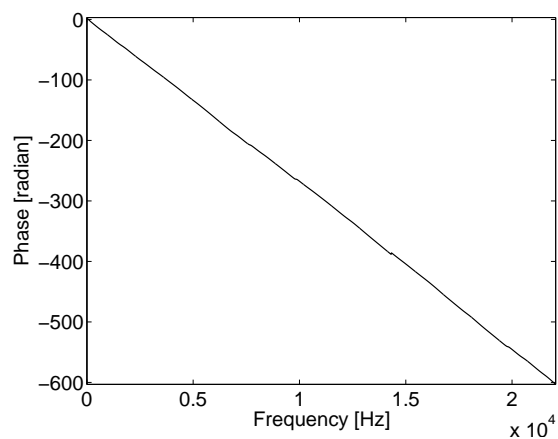


(f) パターン 4 の抽出結果

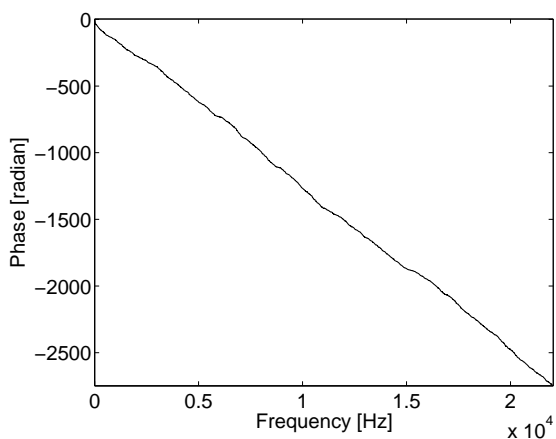
図 A.7: 90 °方向の放射指向特性：周波数応答 (対数スケール)



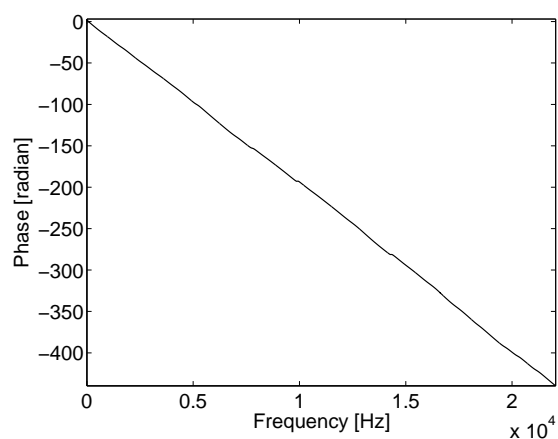
(a) 無響室での測定結果



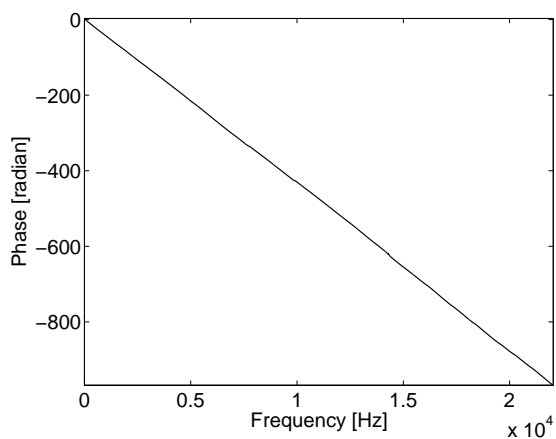
(d) パターン 2 の抽出結果



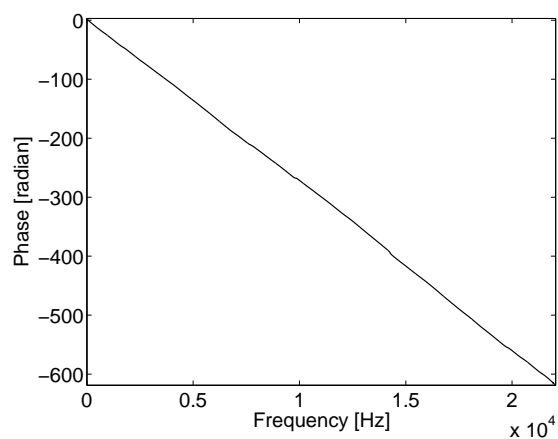
(b) 室内での測定結果 (パターン 1)



(e) パターン 3 の抽出結果

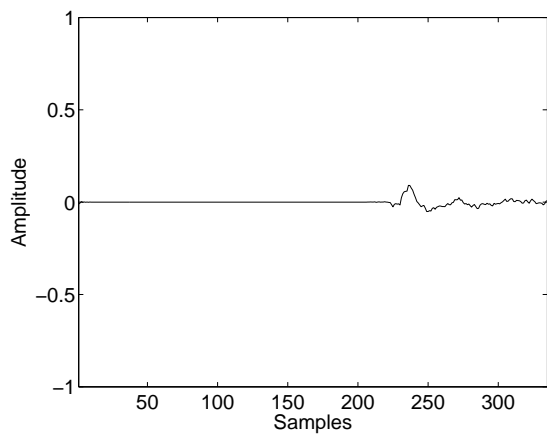


(c) パターン 1 の抽出結果

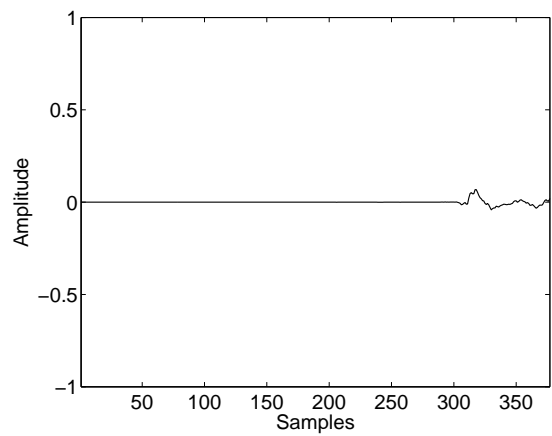


(f) パターン 4 の抽出結果

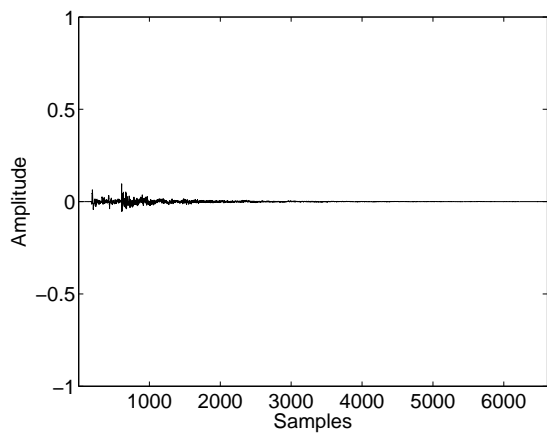
図 A.8: 90 °方向の放射指向特性：位相特性



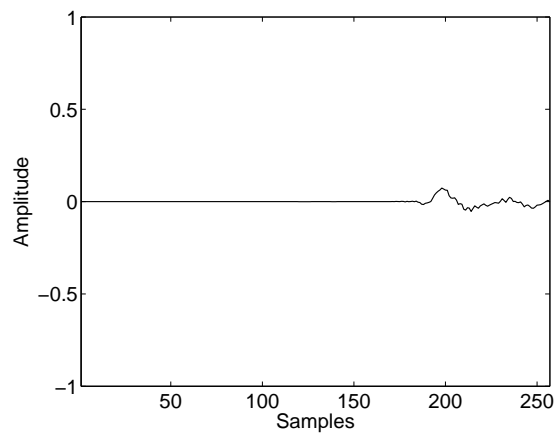
(a) 無響室での測定結果



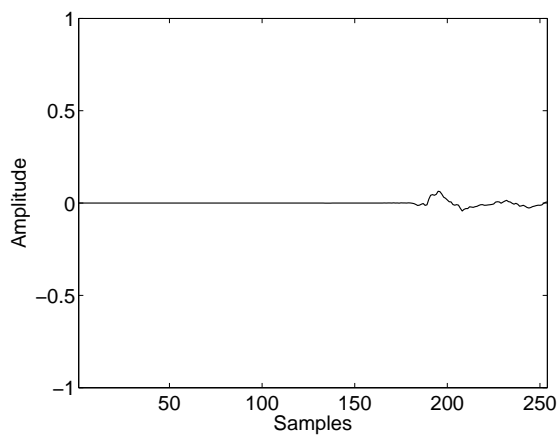
(d) パターン 2 の抽出結果



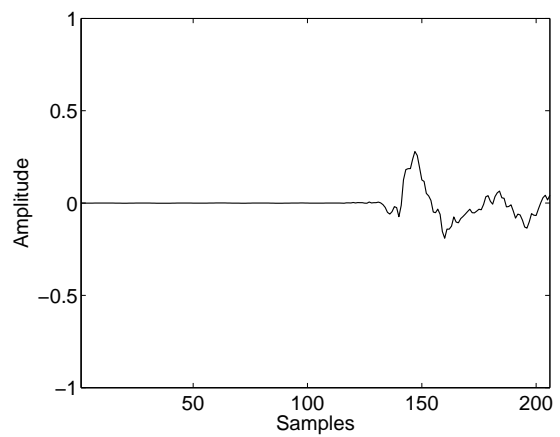
(b) 室内での測定結果 (パターン 1)



(e) パターン 3 の抽出結果

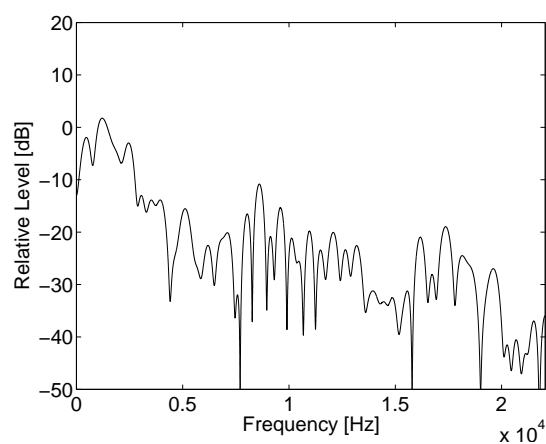


(c) パターン 1 の抽出結果

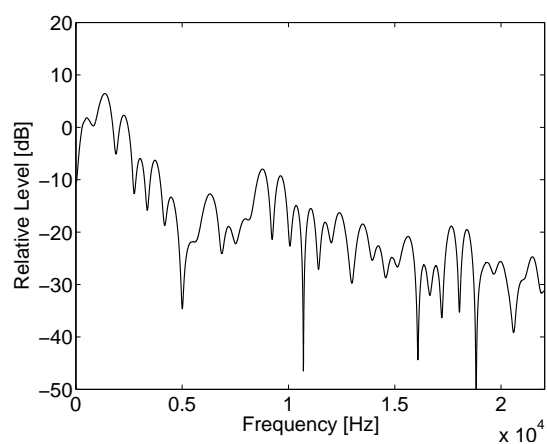


(f) パターン 4 の抽出結果

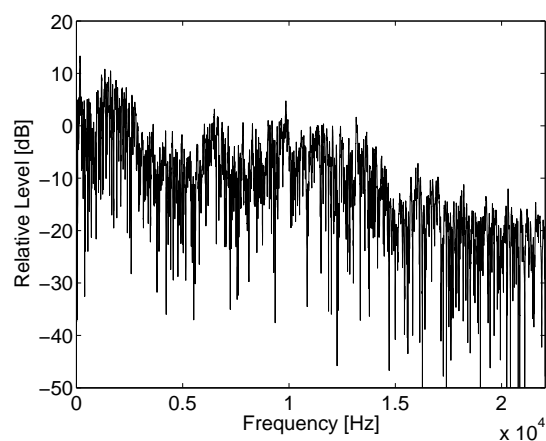
図 A.9: 180 °方向の放射指向特性：インパルス応答



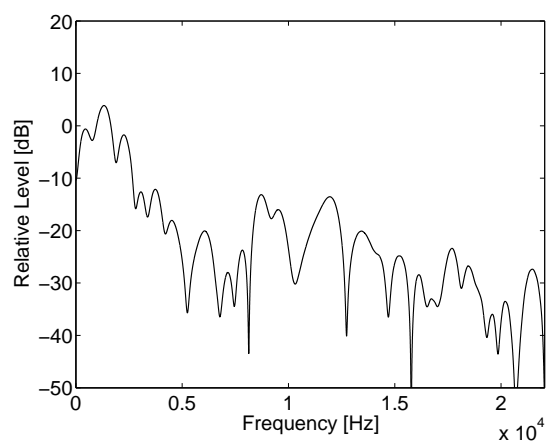
(a) 無響室での測定結果



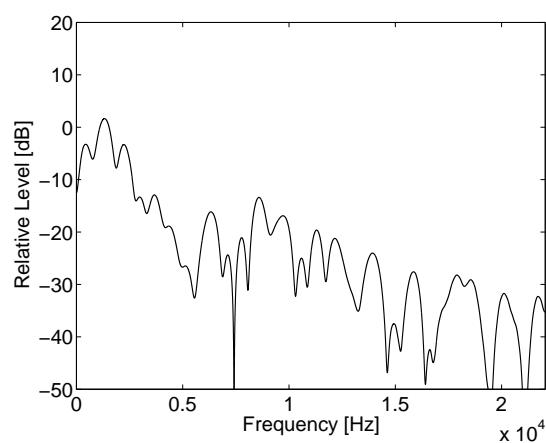
(d) パターン 2 の抽出結果



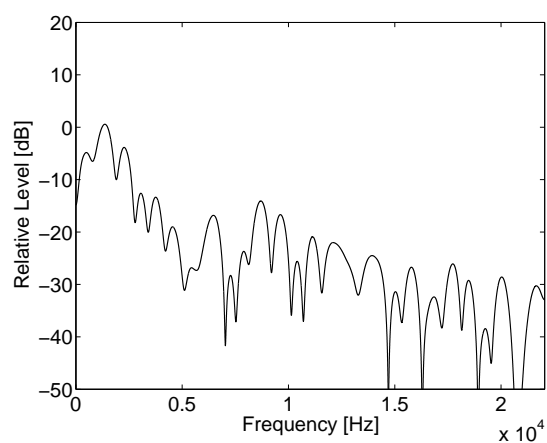
(b) 室内での測定結果 (パターン 1)



(e) パターン 3 の抽出結果

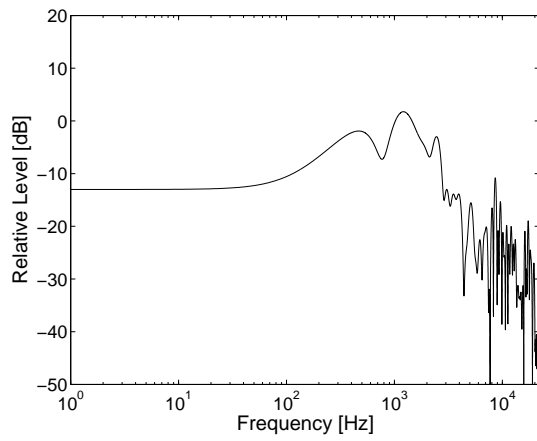


(c) パターン 1 の抽出結果

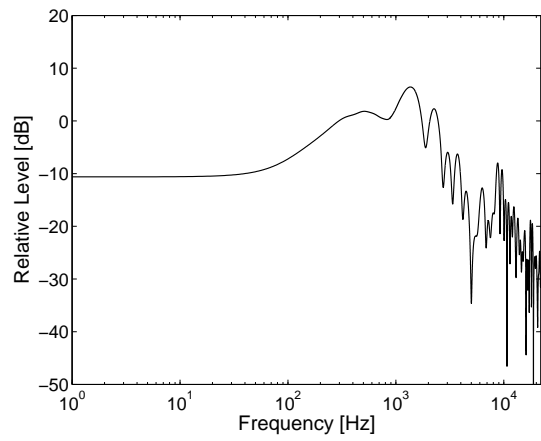


(f) パターン 4 の抽出結果

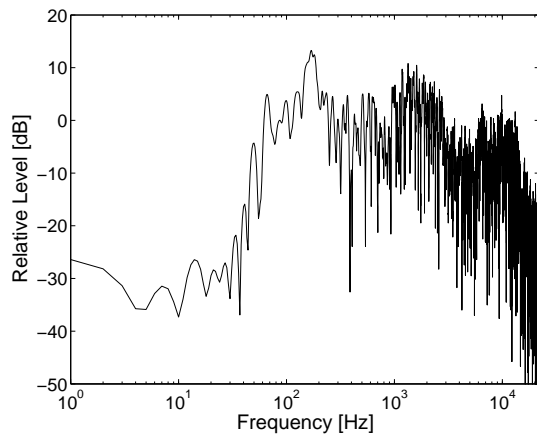
図 A.10: 180 °方向の放射指向特性：周波数応答



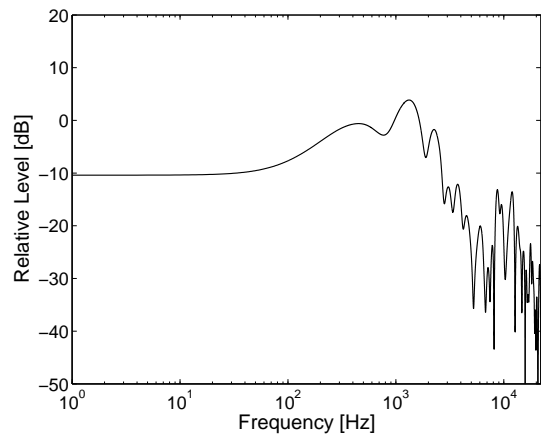
(a) 無響室での測定結果



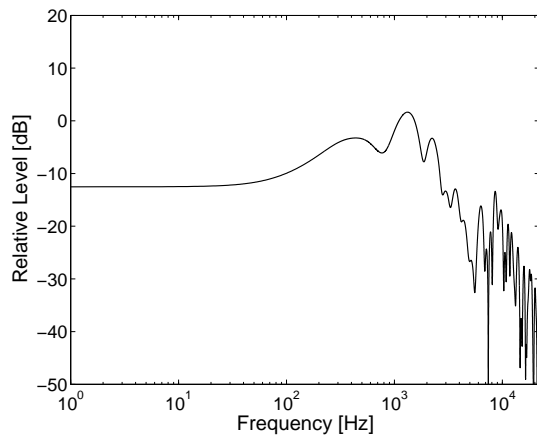
(d) パターン 2 の抽出結果



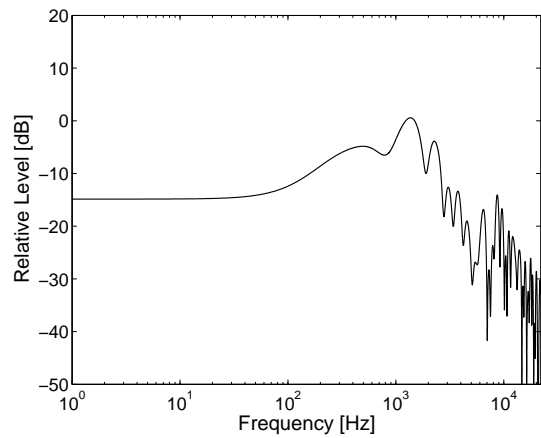
(b) 室内での測定結果 (パターン 1)



(e) パターン 3 の抽出結果

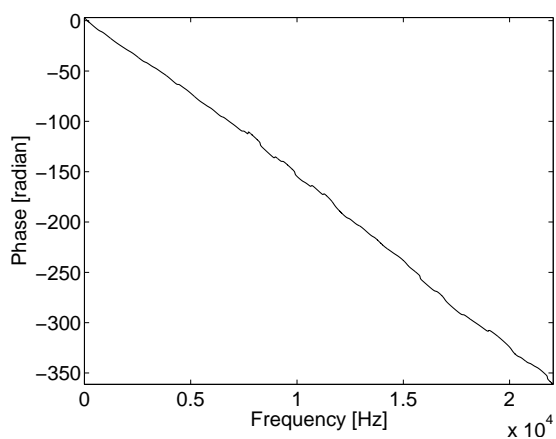


(c) パターン 1 の抽出結果

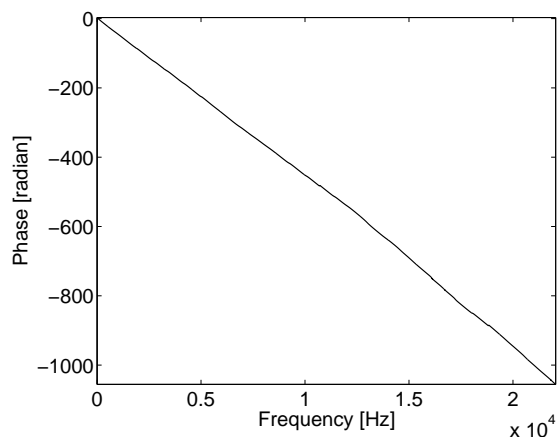


(f) パターン 4 の抽出結果

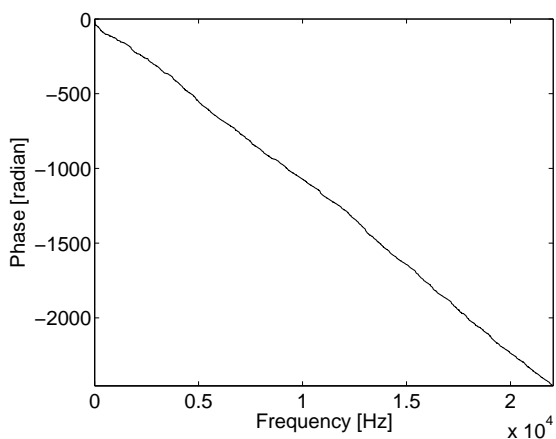
図 A.11: 180 °方向の放射指向特性：周波数応答 (対数スケール)



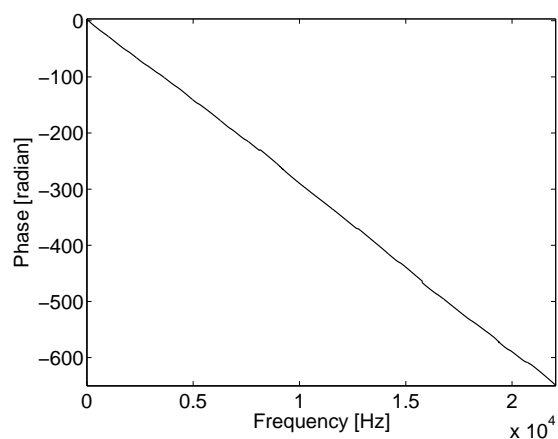
(a) 無響室での測定結果



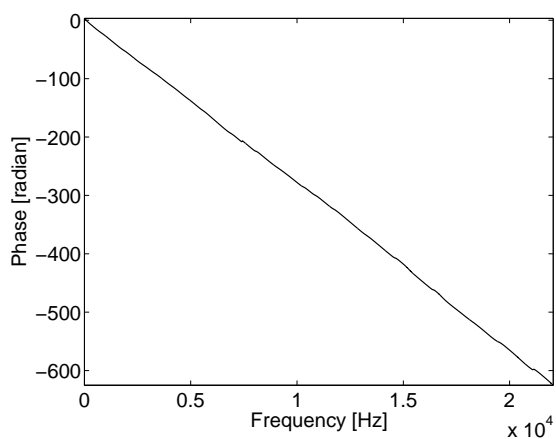
(d) パターン 2 の抽出結果



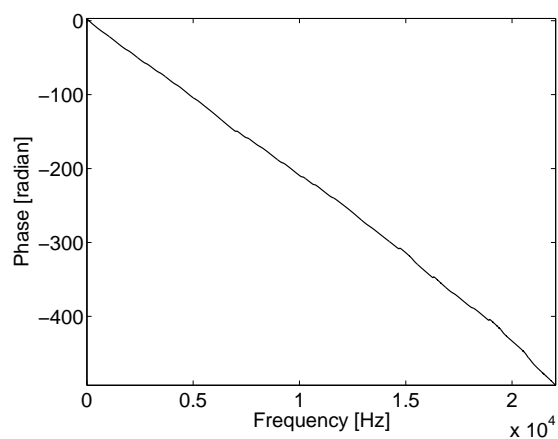
(b) 室内での測定結果 (パターン 1)



(e) パターン 3 の抽出結果



(c) パターン 1 の抽出結果



(f) パターン 4 の抽出結果

図 A.12: 180 °方向の放射指向特性：位相特性



# 謝辞

本論文は，以下に名前を上げることのできなかった方々も含め，多くの方々のご理解やご協力のもとに完成させることができたものである．この場を借り，皆様に感謝の意を表す．

本研究を行うにあたって，東北大学電気通信研究所 鈴木陽一教授には，大学院博士後期課程進学を快く受け入れていただき，本研究に取り組む機会を賜り，研究に取り組む姿勢など，終始にわたりご指導いただいた．また，後期過程1年時には，電気通信研究所一般公開の研究室監事という大役も任せていただき，研究室全体をまとめることを学ぶことができた．ここに深く感謝申し上げる．

本論文を作成するにあたり，東北大学大学院情報科学研究科 青木孝文教授，堀口進教授には，論文の審査において貴重な御意見をいただいた．ここに深く感謝申し上げる．

東北大学電気通信研究所 岩谷幸雄准教授には，後期過程2，3年時に直接のご指導を賜り，東京勤務で大変お忙しい中でも，常に的確なご指導やご助言をいただいた．また，第4章をまとめる上でのNTTの方々との打ち合わせの機会を作っていただき，直接ご同行いただいた．ここに深く感謝申し上げる．情報通信研究機構ユニバーサルメディア研究センター 東北大学電気通信研究所 西村竜一客員准教授には，後期過程1年時に直接のご指導を賜り，また，第4章をまとめる上でも大変貴重なご意見をいただいた．ここに深く感謝申し上げる．

NTT コミュニケーション科学基礎研究所 三好正人博士をはじめ，信号処理研究グループの方々には，第4章をまとめるにあたり研究打ち合わせを快く受け入れていただき，大変貴重なご意見をいただいた．ここに深く感謝申し上げる．

東北大学電気通信研究所 坂本修一助教，宮内良太助教には，大学院ゼミや研究室・グループゼミの場において貴重なご意見や実験機材に関するご指導をいただき，研究室生活においても大変お世話になった．東北大学電気通信研究所 矢入聡助教には，音響測定の実験機材の手配やご指導，ご協力をいただいた．東北大学電気通信研究所技術職員 齊藤文孝氏には，音響測定に関するご指導を賜った．ここに深く感謝申し上げる．

東北大学工学研究科 牧野正三教授，伊藤彰則准教授，伊藤仁助教，東北大学電気通信研究所塩入諭教授，栗木一郎准教授，松宮一道助教，徳島大学 鈴木基之准教授には，大学院ゼミなどの場において，重要なご助言をいただいた．ここに感謝申し上げる．



研究室事務補佐員の小野寺美紀さん，本間理恵さん，科研費技術者の宮内千夏さん，前事務補佐員の小室まりさん，今野亜未さんには，研究室での生活や，研究に専念できるように様々なご配慮をいただいた．ここに感謝する．

研究室の先輩である園田光太郎博士（情報通信研究機構 情報通信セキュリティ研究センター）には，計算機シミュレーションを行う上でワークステーションの有効な利用方法を教えていただいた．ここに感謝する．

研究室の研究員である小林まおり博士には，実験機材に関するご助言や本論文の審査発表練習において貴重なコメントをいただいた．研究室の同輩である魏浩石氏，姜大基氏には，グループ研究や実験設備の構築など様々な場面でお世話になり，また，研究室の学生リーダーとして共に協力し，同期として励ましあいながら研究を行うことができたことは心強かった．後輩である浅川香さんには，論文の日本語校正や研究に専念できるように様々なご協力をいただき，また，同じ後期過程として励ましあいながら研究を行うことができた．後輩である千葉武尊氏，半田浩規氏には，研究室ゼミやグループゼミの場などにおいて貴重なコメントをいただいた．吉田和博氏，小玉純一氏，池田善敬氏には，音響測定の際，ご協力いただいた．増山友介氏，曲谷地哲氏には計算機シミュレーションを行う上で助言をいただいた．その他，研究室の皆様には，研究する上でも生活する上でも大変お世話になり，3年間とても楽しく研究室生活を行うことができた．ここに深く感謝する．

大学に9年間通うことができ，本研究をまとめることができたのは，家族の多大なる支援や理解のおかげであり，また学生寮で9年間安く楽しく生活できたことである．最後に，両親や弟妹，東北大学日就寮に心から感謝する．

## 参考文献

- [1] J. Blauert, *Spatial Hearing*, Cambridge:MIT Press, 1983.
- [2] H. Kuttuff, *Room Acoustics*, Essex, U.K.: Elsevier, 1991.
- [3] L. A. Jeffress, "A place theory of sound localization," *J. Comparative Physiology*, vol. 41, no. 1, pp. 35-39, Feb. 1948.
- [4] M. A. Gerzon, "Periphony: With-height sound reproduction," *J. Audio Eng. Soc.*, vol. 21, no. 1, pp. 2-10, 1973.
- [5] V. Pulkki and T. Lokki, "Creating auditory displays with multiple loudspeakers using VBAP: a case study with DIVA project," *Proc. ICAD 1998*, 1998.
- [6] M. Morimoto and Y. Ando, "On the simulation of sound localization," *J. Acoust. Soc. Jpn. (E)*, vol. 1, no. 2, pp. 167-174, 1980.
- [7] 前迫孝憲, 長谷川保, 清水康敬, "ディジタル処理頭部運動補償によるヘッドホン音像定位の向上", *信学論 (A)*, vol. J73-A, no. 4, pp. 692-699, Apr. 1990.
- [8] J. Kawaura, Y. Suzuki, F. Asano and T. Sone, "Sound localization in headphone reproduction by simulating transfer function from the sound source to the external ear," *J. Acoust. Soc. Jpn. (E)*, vol. 12, no. 5, pp. 203-216, 1991.
- [9] J.W. Scarpaci, H.S. Colburn and J.A. White, "A system for real-time virtual auditory space," *Proc. ICAD 2005*, 2005.
- [10] D.S. Brungart, B.D. Simpson and A.J. Kordik, "The detectability of headtracker latency in virtual audio displays," *Proc. ICAD 2005*, 2005.
- [11] 矢入聡, 岩谷幸雄, 鈴木陽一, "頭部運動感応型ソフトウェア聴覚ディスプレイの開発", *日本バーチャルリアリティ学会論文誌*, vol.11, no. 3, pp. 437-446, Sep. 2006.
- [12] Y. Iwaya and Y. Suzuki, "Rendering moving sound with the doppler effect in sound space," *Applied Acoustics*, vol. 68, no. 8, pp. 916-922, Aug. 2007.

- [13] M. Otani and T. Hirahara, "Auditory artifacts due to switching head-related transfer functions of a dynamic virtual auditory display," *IEICE Trans. Fundamentals*, vol. E91-A, no. 6, pp. 1320-1328, June 2008.
- [14] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am.*, vol. 65, no. 4, pp. 943-950, Apr. 1979.
- [15] V. Mellert, "Construction of a dummy head after new measurements of the threshold of hearing," *J. Acoust. Soc. Am.*, vol.51, no. 4, pp.1359-1361, Apr. 1972.
- [16] H. Møller, "Fundamentals of binaural technology," *Applied Acoustics*, vol.36, no, 3-4, pp.171-218, 1992.
- [17] D. H. Cooper and J. L. Bauck, "Prospects for transaural recording," *J. Audio Eng. Soc.*, vol.37, no. 1/2, pp.3-19, Feb. 1989.
- [18] I. Toshima, H.Uematsu and T.Hirahara, "A steerable dummy head that tracks three-dimensional head movement: *TeleHead*," *J. Acoust. Sci & Tech.*, vol.24, no.5, pp.327-329, 2003.
- [19] H. Fletcher, "Symposium on wire transmission of symphonic music and its reproduction on auditory perspective: Basic requirement," *Bell System Technical Journal*, vol.13, no.2, pp.239-244, 1934.
- [20] M. Camras, "Approach to recreating a sound field," *J. Acoust. Soc. Am.*, vol.43, no.6, pp.1425-1431, June 1968.
- [21] A. J. Berkhout, D. de Vries and P. Vogel, "Acoustic control by wave field synthesis," *J. Acoust. Soc. Am.*, vol.93, no.5, pp.2764-2778, May 1993.
- [22] P. M. Morse and K. U. Ingard, "*Theoretical acoustics*," Princeton University Press, 1968.
- [23] S. Ise, "A principle of sound field control based on the Kirchhoff-Helmholtz integral equation and the theory of inverse systems," *ACUSTICA - Acta Acustica*, vol.85, no.1, pp.78-87, Jan. 1999.
- [24] S. Takane, Y. Suzuki, and T. Sone, "A new method for global sound reproduction based on Kirchhoff's integral equation," *ACUSTICA -Acta Acustica*, vol.85, no.2, pp.250-257, Mar. 1999.

- [25] S. Ise, “The development of the sound field sharing system based on the boundary surface control principle,” *Proc. ICA 2007*, ELE-04-003, 2007.
- [26] S. Enomoto, Y. Ikeda, S. Nakamura and S. Ise, “Three-dimensional sound field reproduction and recording system based on boundary surface control principle,” *Proc. ICAD 2008*, 2008.
- [27] N. Suzuki, I. Umata, T. Kitamura, H. Ando and N. Inoue, “Toward adaptive interaction - The effect of ambient sounds in an ultra-realistic communication system,” *Human Interface and the Management of Information. Interacting in Information Environments*, pp. 185-194, Aug. 2007.
- [28] H. K. Dunn and D. W. Farnsworth, “Exploration of pressure field around the human head during speech,” *J. Acoust. Soc. Am.*, vol. 10, no. 3, pp. 184-199, Jan. 1939.
- [29] G. A. Studebaker, “Directivity of the human vocal source in the horizontal plane,” *Ear and Hearing*, vol. 6, no. 6, pp. 315-319, Nov. 1985.
- [30] W. T. Chu and A. C. Warnock, “Detailed directivity of sound fields around human talkers,” *Tech. Rep. IRC-RR-144*, Insititute for Research in Construction, National Research Council Canada, 2002.
- [31] K. Niwa, T. Nishino and K. Takeda, “Encoding large array signals into a 3D sound field representation for selective listening point audio based on blind source separation,” *Proc. of ICASSP 2008*, pp. 181-184, Apr. 2008.
- [32] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*, Berlin: Springer-Verlag, 2001.
- [33] J. Benesty, J. Chen and Y. Huang, *Microphone Arrays Signal Processing*, Berlin: Springer-Verlag, 2008.
- [34] K. Abed-Meraim, W. Qiu and Y. Hua, “Blind system identification,” *Proc. IEEE*, vol. 85, no. 8, pp. 1310-1322, Aug. 1997.
- [35] 勝本道哲, 山肩洋子, 木村敏幸, “球形スピーカによる異なる放射特性を持つ音響表現の実現”, 信学技報, EA2008-04, pp. 23-26, Apr. 2008.

- [36] *IEEE Standard for a Smart Transducer Interface for Sensors and Actuators - Transducer to Microprocessor Communication Protocols and Transducer Electric Data Sheet (TEDS) Format*, IEEE Std. 1451.2-1997, Sep. 1997.
- [37] K. Lee, "IEEE 1451: A standard in support of smart transducer networking," *Proc. of 17th IEEE Instrum. Meas. Technol. Conf.*, vol. 2, pp. 525-528, May 2000.
- [38] *IEEE Standard for a Smart Transducer Interface for Sensors and Actuators - Mixed-Mode Communication Protocols and Transducer Electric Data Sheet (TEDS) Format*, IEEE Std. 1451.4-2004, May 2004.
- [39] R. C. Cabot, "Measuring AES-EBU digital audio interface," *Proc. AES 87th Convention*, Oct. 1989.
- [40] A. Freed and D. Wessel, "Communication of musical gesture using the AES / EBU digital audio standard," *Proc. of Int. Computer Music Conf.*, pp. 220-223, Oct. 1998.
- [41] N. Aoshima, "Computer-generated pulse signal applied for sound measurement," *J. Acoust. Soc. Am.*, vol. 69, no. 5, pp. 1484-1488, May 1981.
- [42] Y. Suzuki, F. Asano, H. Y. Kim and T. Sone, "An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses," *J. Acoust. Soc. Am.*, vol. 97, no. 2, pp. 1119-1123, Feb. 1995.
- [43] 鈴木陽一, 浅野太, 金学胤, 曾根敏夫, "時間引き延ばしパルスの設計法に関する考察", *信学技報*, EA92-86, pp. 17-24, 1992.
- [44] T. Shan, M. Wax, and T. Kailath, "On Spatial Smoothing for Direction-of-Arrival Estimation of Coherent Signals," *IEEE Trans. Acoust., Speech & Signal Process.*, vol. 33, no. 4, pp. 806-811, Aug. 1985.
- [45] R. O. Schmidt, "Multiple Emitter Location and Signal Parameter Estimation," *IEEE Trans. Antennas, Propag.*, vol. 34, no. 3, pp. 276-280, Mar. 1986.
- [46] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 24, no. 4, pp. 320-327, Aug. 1976.
- [47] S. A. Schelkunoff, "A mathematical theory of linear arrays," *Bell Syst. Tech. J.*, vol. 22, pp. 80-107, Jan. 1943.

- [48] D. E. Dudgeon, "Fundamentals of digital array processing," *Proc. IEEE*, vol. 65, no. 6, pp. 898-904, June 1977.
- [49] J. L. Flanagan, J. D. Johnson, R. Zahn and G. W. Elko, "Computer-steered microphone arrays for sound transduction in large room," *J. Acoust. Soc. Am.*, vol. 78, no. 5, pp. 1508-1518, Nov. 1985.
- [50] 大賀寿郎, 山崎芳男, 金田豊, 音響システムとディジタル処理, 電子情報通信学会, 1995.
- [51] S. U. Pillai, *Array signal processing*, New-York: Springer-Verlag, 1989.
- [52] H. Nomura, Y. Kaneda and J. Kojima, "Optimum gains of a delay-and-sum microphone array for near field," *J. Acoust. Soc. Am.*, vol. 100, no. 4, pp. 2697, Oct. 1996.
- [53] 野村博昭, 金田豊, 小島順治, "近接音場型マイクロホンアレー", 日本音響学会誌, vol. 53, no. 2, pp. 110-116, 1997.
- [54] R. A. Kennedy, T. Abhayapala and D. B. Ward, "Broadband nearfield beamforming using a radial beampattern transformation," *IEEE Trans. Signal Process.*, vol. 46, no. 8, pp. 2147-2156, Aug. 1998.
- [55] J. G. Ryan and R. A. Goubran, "Array optimization applied in the near field of a microphone array," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 8, no. 2, pp. 173-176, Mar. 2000.
- [56] 菊間信良, アレーアンテナによる適応信号処理, 科学技術出版, 1998.
- [57] R. T. Williams, S. Prasad, A. K. Mahalanabis and L. H. Sibul, "An improved spatial smoothing technique for estimation in a multipath environment," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 36, no. 4, pp. 425-432, Apr. 1988.
- [58] S. U. Pillai and B. H. Kwon, "Forward/backward spatial smoothing techniques for coherent signal identification," *IEEE Trans. Acoust., Speech & Signal Process.*, vol. 37, no. 1, pp. 8-15, Jan. 1989.
- [59] B. D. Rao and K. V. S. Hari, "Weighted subspace methods and spatial smoothing: analysis and comparison," *IEEE Trans. Signal Process.*, vol. 41, no. 2, pp. 788-803, Feb. 1993.

- [60] C. Qi, Y. Wang, Y. Zhang and Y. Han, "Spatial difference smoothing for DOA estimation of coherent signals," *IEEE Trans. Signal Process. Lett.*, vol. 12, no. 11, pp. 800-802, Nov. 2005.
- [61] R. ツルミュール著, 瀬川富士, 高市成方共訳. マトリクスの理論と応用 - 理工学への応用の技法 - , ブレイン図書出版, 1972.
- [62] B. Friedlander and A.J. Weiss, "Direction finding using spatial smoothing with interpolated arrays," *IEEE Trans. Aerosp., Electron.*, vol. 28, no. 2, pp. 574-587, Apr. 1992.
- [63] A.J. Weiss and B. Friedlander, "Performance analysis of spatial smoothig with interpolated arrays," *IEEE Trans. Signal Process.*, vol. 41, no. 5, pp. 1881-1892, May 1993.
- [64] M. Wax and J. Sheinvaïd, "Direction finding of coherent signals via spatial smoothing for uninform circular arrays," *IEEE Trans. Antennas, Propag.*, vol. 42, no. 5, pp. 613-620, May 1994.
- [65] 岡村敦, 米澤ルミ子, 桐本哲朗, "アレー補間処理に窓関数を導入した空間スムージング型超解像測角法", 信学論 (B), vol. J82-B, no. 6, pp. 1185-1192, June 1999.
- [66] 岡村敦, 藤坂貴彦, "仮想入射波を加える超分解能測角法", 1999 信学総大, B1-1, 1999.
- [67] 畝田道雄, 福江敏彦, 外園博一, "AVW 法と段階的補間 MUSIC 法を組み合わせた不等間隔配列リニアアレーによるコヒーレント波の到来方向推定", 信学論 (B), vol. J84-B, no. 12, pp. 2342-2350, Dec. 2001.
- [68] Y. R. Zheng, R. A. Goubran, M. El-Tanany and H. Shi, "A microphone array system for multimedia application with nearfield signal target," *IEEE Sens. J.*, vol. 5, no. 6, pp. 1395-1406, Dec. 2005.
- [69] 小貫輝義, 齋藤文孝, 鈴木陽一, "無指向性 32 面体スピーカの紹介", 電気関係東北支部連合大会講演論文集, pp. 220-220, Aug. 2005.
- [70] M. Delcroix, T. Hikichi and M. Miyoshi, "Blind dereverberation algorithm for speech signals based multi-channel linear prediction," *Acoust. Sci. & Tech.*, vol. 26, no. 5, pp. 432-439, Sep. 2005.

- [71] M. Delcroix, T. Hikichi and M. Miyoshi, "Precise dereverberation using multi-channel linear prediction," *IEEE Trans. Audio Speech Lang. process.*, vol. 15, no. 2, pp. 430-440, Feb. 2007.
- [72] S. T. Neely and J. B. Allen, "Invertibility of a room impulse response," *J. Acoust. Soc. Am.*, vol. 66, no. 1, pp. 165-169, July 1979.
- [73] M. Miyoshi and Y. Kaneda, "Inverse filtering of room acoustics," *IEEE Trans. Speech, Signal Process.*, vol. 36, no. 2, pp. 145-152, Feb. 1988.
- [74] B. Widrow and E. Walach, "Adaptive signal processing for adaptive control," *Proc. ICASSP 84*, vol. 9, pp. 191-194, Mar. 1984.
- [75] B. Widrow and E. Walach, *Adaptive inverse control*, Prentice Hall, 1995.
- [76] L. A. Poole, G. E. Warnaka and R. C. Cutter, "The implementation of digital filters using a modified Window-Hoff algorithm for the adaptive cancellation," *Proc. ICASSP 84*, vol. 9, pp. 215-218, Mar. 1984.
- [77] R. W. Scott and B. D. O. Anderson, "Least order, stable solutions of exact model matching problem," *Automatica*, vol. 14, pp. 481-492, 1978.
- [78] Y. Sato, "A method of self-recovering equalization for multichannel amplitude-modulation," *IEEE Trans. Commun.*, vol. COM-23, pp. 679-682, June 1975.
- [79] D. N. Godard, "Self-recovering equalization and carrier tracking in a two-dimensional data communication system," *IEEE Trans. Commun.*, vol. COM-28, pp. 1867-1875, 1980.
- [80] D. Hatzinakos and C. L. Nikias, "Blind equalization using a trispectrum based algorithm," *IEEE Trans. Commun.*, vol. COM-39, pp. 669-682, 1991.
- [81] G. Xu, H. Liu, L. Tong and T. Kailath, "A least-square approach to blind channel identification," *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2982-2993, Dec. 1995.
- [82] L. Tong, G. Xu, B. Hassihi and T. Kailath, "Blind identification and equalization based on second-order statistics: a frequency domain approach," *IEEE Trans. Inform. Theory*, vol. 41, no. 1, pp. 329-334, Jan. 1995.



- [83] H. H. Zeng and L. Tong, "Blind channel estimation using the second-order statistics: algorithms," *IEEE Trans. Signal Process.*, vol. 45, pp. 1919-1930, Aug. 1997.
- [84] H. Liu, G. Xu, L. Tong and T. Kailath, "Recent developments in blind channel equalization: from cyclostationarity to subspaces," *Signal Process.*, vol. 50, no. 1-2, pp. 83-99, Apr. 1996.
- [85] S. Haykin, *Adaptive Filter Theory*, 3rd ed., Prentice Hall, 1996.
- [86] Y. Huang and J. Benesty, "Adaptive multi-channel least mean square and Newton algorithms for blind channel identification" *Signal Process.*, vol. 82, no. 8, pp. 1127-1138, Aug. 2002.
- [87] Y. Huang J. Benesty and J. Chen, "Optimal step size of the adaptive multichannel LMS algorithm for blind SIMO identification," *IEEE Signal Process. Lett.*, vol. 12, no. 3, pp. 173-176, Mar. 2005.
- [88] Y. Huang and J. Benesty, "A class of frequency-domain adaptive approaches to blind multichannel identification," *IEEE Trans. Signal Process.*, vol. 51, no. 1, pp. 11-24, Jan. 2003.
- [89] Z. L. Yu and M. H. Er, "A robust frequency domain adaptive blind multichannel identification algorithm for acoustic applications," *Proc. ICASSP2004*, vol. 2, pp. 25-28, May 2004.
- [90] R. Ahmad, A. W. H. Andy and P. A. Naylor, "A practical adaptive blind multichannel estimation algorithm with application to acoustic impulse responses," *Proc 15th International Conference on Digital Signal Process. 2007*, pp. 31-34, July 2007.
- [91] M. A. Haque and M. K. Hasan, "Variable step-size multichannel frequency-domain LMS algorithm for blind identification of finite impulse response systems," *IEEE Signal Process.*, vol. 1, no. 4, pp. 182-189, Dec. 2007.
- [92] E. Moulines, P. Duhamel, J.-F. Cardoso and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 516-525, Feb. 1995.

- [93] K. Furuya and Y. Kaneda, "A two-channel blind deconvolution of nonminimum phase FIR systems," *IEICE Trans. Fundamentals*, vol. E80-A, no. 5, pp. 804-808, May 1997.
- [94] S. Gannot and M. Moonen, "Subspace methods for multimicrophone speech dereverberation," *EURASIP J. Appl. Signal Process.*, vol. 2003, no. 11, pp. 1074-1090, 2003.
- [95] P. P. Vaidyanathan, *Multirate System and filterbanks*, Prentice Hall, 1993.
- [96] A. Gilloire and M. Vetterli, "Adaptive filtering in subbands with critical sampling: analysis, experiments, and application to acoustic echo cancellation," *IEEE Trans. Signal Process.*, vol. 40, no. 8, pp. 1862-1875, Aug. 1992.
- [97] S. Weiss, M. Hartneck and R. W. Stewart, "Design of near perfect reconstruction oversampled filterbanks for subband adaptive filters," *IEEE Trans. Circuits Syst. II*, vol. 46, no. 8, pp. 1081-1085, Aug. 1999.
- [98] J. P. Reilly, M. Willbur, M. Seibert and N. Ahmadvand, "The complex subband decomposition and its application to the decimation of large adaptive filtering problem," *IEEE Trans. Signal Process.*, vol. 50, no. 11, pp. 2730-2743, Nov. 2002.
- [99] M. J. Daly and J. R. Reilly, "Blind deconvolution using Bayesian methods with application to the dereverberation of speech," *Proc. ICASSP 2004*, vol. 2, pp. 1009-1012, May 2004.
- [100] P. Common, "Independent component analysis, a new concept?," *Signal Process.*, vol. 36, pp. 287-314, 1994.
- [101] T. W. Lee, *Independent Component Analysis*, Kluwer Academic Publishers, 1998.
- [102] Y. Zhou, and B. Xu, "Blind source separation in frequency domain," *Signal Processing*, vol. 83, no. 9, pp. 2037-2046, Sep. 2003.
- [103] 古家賢一, 片岡章俊, "チャンネル間相関行列と音声の白色化フィルタを用いた Semi-blind 残響抑圧", *信学論 (A)*, vol. J88-A, no. 10, pp. 1089-1099, Oct. 2005.
- [104] K. Furuya and A. Kataoka, "Robust speech dereverberation using multichannel blind deconvolution with spectral subtraction," *IEEE Trans. Audio, Speech, Lang. process.*, vol. 15, no. 5, pp. 1579-1591, July 2007.

- [105] S. Amari, S. C. Douglas, A. Cichocki and H. H. Yahg, "Multichannel blind deconvolution and equalization using the natural gradient," *Proc. IEEE Workshop on Signal Procssing in Advances in Wireless Communications*, Paris, pp.101-104, April 1997.
- [106] X. Sun and S. C. Douglas, "A natural gradient convolutive blind source separation algorithm for speech mixtures," *Proc. ICA 2001*, pp.59-64, 2001.
- [107] R. Aichner, S. Araki and S. Makino, "Time domain blind source separation of non-stationary convolved signals by utilizing geometric beamforming," *Proc. NNSP 2002*, pp.445-454, 2002.
- [108] M. Miyoshi, "Estimating AR parameter-sets for linear-recurrent signals in convolutive mixtures," *Proc. ICA 2003*, pp. 585-589, Apr. 2003.
- [109] T. Hikichi and M. Miyoshi, "Blind algorithm for calculating the common poles based on linear prediction," *Proc. ICASSP 2004*, vol. 4, pp. 89-92, May 2004.
- [110] M. Miyoshi, M. Delcroix and K. Kinoshita, "Calculating inverse filters for speech dereverberation," *IEICE Trans. Fundermentals*, vol. E91-A, no. 6, pp. 1303-1309, June 2008.
- [111] D. A. Harville, *Matrix Algebra From a Statistician's Perspective*, NewYork, Springer-Verlag, 1997.
- [112] S. Rombouts and K. heybe, "An accurate and efficient algorithm for the computation of the charecteristic polynomial of a general square matrix," *J. Comput. Phys.*, vol.140, pp.453-458, 1998.
- [113] T. Hikichi, M. Delcroix and M. Miyoshi, "Speech dereverberation algorithm using transfer function estimates with overestimated order," *Acoust. Sci. & Tech.*, vol. 27, no. 1, pp. 28-35, 2006.
- [114] N. D. Gaubitch, P. A. Naylor and D. B. Ward, "On the use of linear prediction for dereverberation of speech," *Proc. Int. Workshop Acoust. Echo Noise Control*, vol.1, pp.99-102, 2003.
- [115] L. Ljung, *System Identification: Theory for the User*, Prentice Hall, 1987.

- [116] K. Kinoshita, M. Delcroix, T. Nakatani and M. Miyoshi, “A linear prediction-based microphone array for speech dereverberation in a realistic sound field,” *Proc. AES 13th Regional Convention*, July 2007.
- [117] 後藤真孝, 橋口博樹, 西村拓一, 岡隆一, “RWC 研究用音楽データベース: クラシック音楽データベースとジャズ音楽データベース”, 情報処理学会音楽情報処理研究会研究報告, vol. 2002-MUS-44-5, pp. 25-32, 2002.
- [118] J. G. Neuhoff, “Twist and shout: Audible facing angles and dynamic rotation,” *Ecological Psychology*, vol. 15, no. 4, pp. 335-351, Jan. 2003.
- [119] 高野洋樹, 刈穂治英, 島田正治, 杉山精, “話者の発話方向知覚に関する一検討”, 信学技報 EA2005-63, pp. 37-42, 2005.
- [120] H. Kato, H. Takemoto and R. Nishimura, “Reception of speaker’s facing angle,” *J. Acoust. Soc. Am.*, vol. 123, no. 5, pp. 3294, May 2008.
- [121] K. Nakadai, H. Nakajima, K. Yamada, Y. Hasegawa, T. Nakamura and H. Tsujino, “Sound source tracking with directivity pattern estimation using a 64 ch microphone array,” *IROS 2005*, pp. 1690-1696, Aug. 2005.
- [122] T. Hikichi, M. Delcroix and M. Miyoshi, “Inverse filtering for speech dereverberation less sensitive to noise and room transfer fluctuations,” *EURASIP J. Advances in Signal Process.*, Article ID 34013, 2007.
- [123] Y. Huang, J. Benesty and J. Chen, “A blind channel identification-based two-stage approach to separation and dereverberation of speech signals in a reverberant environment,” *IEEE Trans. Speech, Audio Process.*, vol. 13, no. 2, pp.882-895, Sep. 2005.
- [124] Y. Huang, J. Benesty and J. Chen, “Identification of acoustic MIMO systems: Challenges and opportunities,” *Signal Process.*, vol.86, no.6, pp.1278-1295, June 2006.
- [125] J. Benesty, J. Chen and Y. Huang, “On microphone-array beamforming a MIMO acoustic signal processing perspective,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 3, pp. 1053-1065, Mar. 2007.